

CENTRO UNIVERSITÁRIO FEI

GABRIEL LUCAS FACCIN

IGNACIO CITTA

LEONARDO FORMENTI ANTONIO

MARCELLA MONTEIRO DA SILVA

**PREVISÃO DE DEMANDA ATRAVÉS DA UTILIZAÇÃO DE FERRAMENTAS DE  
INTELIGÊNCIA ARTIFICIAL**

São Bernardo do Campo

2020

GABRIEL LUCAS FACCIN  
IGNACIO CITTA  
LEONARDO FORMENTI ANTONIO  
MARCELLA MONTEIRO DA SILVA

**PREVISÃO DE DEMANDA ATRAVÉS DA UTILIZAÇÃO DE FERRAMENTAS DE  
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado ao Centro Universitário FEI, como parte dos requisitos necessários para obtenção do título de Bacharel em Engenharia de Produção. Orientado pelo Prof. Dr. Fábio Lima.

São Bernardo do Campo

2020

PREVISÃO DE DEMANDA ATRAVÉS DA UTILIZAÇÃO DE  
FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL / Gabriel Lucas  
Faccin...[et al.]. São Bernardo do Campo, 2020.  
91 p. : il.

Trabalho de Conclusão de Curso - Centro Universitário FEI.  
Orientador: Prof. Dr. Fábio Lima.

1. Previsão de demanda. 2. Inteligência Artificial. 3. Machine  
Learning. 4. Deep Learning. I. Lucas Faccin, Gabriel. II. Citta,  
Ignacio. III. Formenti Antonio, Leonardo. IV. Monteiro da Silva,  
Marcella. V. Lima, Fábio, orient. VI. Título.

GABRIEL LUCAS FACCIN  
IGNACIO CITTA  
LEONARDO FORMENTI ANTONIO  
MARCELLA MONTEIRO DA SILVA

**PREVISÃO DE DEMANDA ATRAVÉS DA UTILIZAÇÃO DE FERRAMENTAS DE  
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso,  
apresentado ao Centro Universitário FEI,  
como parte dos requisitos necessários  
para obtenção do título de Bacharel em  
Engenharia de Produção.

Comissão julgadora

---

Prof. Dr. Fábio Lima

---

Prof. Dr. Alexandre Augusto Massote

---

Prof. Dr. José Agostinho Baitello

São Bernardo do Campo

18/06/2020

Dedicamos este trabalho aos nossos pais, orientador e, principalmente, a todos os professores do Centro Universitário FEI que fizeram parte de nossa trajetória e que nos capacitaram ao decorrer dos últimos anos, permitindo que nós crescêssemos não só como pessoas, mas principalmente como profissionais da área da Engenharia.

## **AGRADECIMENTOS**

Agradecemos primeiramente nossas famílias por toda dedicação, paciência e suporte ao longo dos anos, contribuindo diretamente para que nós pudéssemos desfrutar de uma caminhada longa e enriquecedora.

Ao nosso orientador Prof. Dr. Fábio Lima por não só ter ministrado as aulas com maestria, mas como também por ter se disposto a nos auxiliar ao longo da elaboração do presente trabalho e, principalmente, por ter nos incentivado a realizar as pesquisas neste campo de estudo.

Aos professores do Centro Universitário FEI que fizeram parte de toda nossa jornada no curso de Engenharia de Produção, mostrando-se sempre dispostos a ajudar e contribuir com nosso aprendizado.

Agradecemos também, ao Sr. José Eduardo Antonio por ter disponibilizado para nós os dados de sua companhia, permitindo que os conceitos abordados ao longo deste trabalho fossem aplicados dentro de sua empresa.

“Lembre-se que as pessoas podem tirar tudo de você, menos o seu conhecimento.”  
– Albert Einstein

## RESUMO

Com o intuito de buscar cada vez mais vantagens competitivas frente ao mercado, as organizações vêm buscando ferramentas gerenciais, e até mesmo mudanças em seus processos, para otimizar suas produções. A projeção de demanda, por exemplo, é uma das ferramentas utilizadas para diminuir o estoque excessivo e, conseqüentemente, diminuir o capital de giro imobilizado. No entanto, as previsões estão sujeitas a erros e é de extrema importância escolher o método mais adequado de acordo com as características dos dados da organização. Para a realização do presente trabalho, será utilizada uma base de dados com variados SKUs (*Stock Keeping Units*) de uma empresa paulista de componentes eletrônicos. Diversos métodos são tradicionalmente utilizados para a projeção como por exemplo: Séries temporais, ARIMA, SARIMA, Suavização Exponencial Simples, Holt Winter, Croston, etc e, recentemente, métodos baseados em aprendizagem de máquina têm ganhado força também através da utilização das redes neurais. O presente trabalho utiliza as técnicas mencionadas acima com o intuito Este trabalho utiliza as técnicas mencionadas acima com o intuito de verificar qual modelo mais adequado para a projeção de demanda dos componentes eletrônicos da empresa estudada.

Palavras-chave: Previsão de Demanda. Inteligência Artificial. *Deep Learning*.

## ABSTRACT

With the purpose of achieving more and more competitive vantages in order to satisfy a growing market, organizations have been looking to management tools and even considering changes in their processes in order to optimize production. One of these tools is the demand forecasting, which is utilized to reduce the overstock and consequently decreasing the fixed working capital. However, these forecasts are vulnerable to errors and it is extremely important that the most suitable method is chosen according to the characteristics of the organization data. To carry out the present work, a database containing several SKUs (Stock Keeping Units) of an electronic components company based in Mogi Mirim. Various methods are traditionally utilized for the projection, such as time series, ARIMA, SARIMA, Simple exponential smoothing, Holt Winter, Croston, etc and more recently, methods based on machine learning have been gaining strength throught the utilization of neural networks. The present work utilizes the techniques mentioned above with the purpose of asserting which model is most suitable for the demand forecasting of the electronic components company studied.

Keywords: Demand Forecasting. Artificial Intelligence. *Deep Learning*.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Comportamento de demandas.....	25
Figura 2 - Curva ABC .....	39
Figura 3 - Colossus, máquina desenvolvida por Alan Turing .....	41
Figura 4 - Representação gráfica do Teste de Turing .....	42
Figura 5 - Esquemática de um neurônio biológico.....	45
Figura 6 - O Neurônio Booleano de McCulloch e implementações de algumas funções booleanas. ....	47
Figura 7 - Funções lineares ou não .....	49
<i>Figura 8 - Funções Linearmente Separáveis</i> .....	49
Figura 9 - Lógica XOR.....	50
Figura 10 - Modelo de Rede Neural Artificial.....	50
Figura 11 - Rede Neural com função de ativação .....	51
Figura 12 - Gráfico Função Degrau .....	52
Figura 13 - Gráfico Função Degrau Bipolar.....	53
Figura 14 - Gráfico Função Rampa Simétrica .....	53
Figura 15 - Função de ativação Sigmoidal .....	54
Figura 16 - Função de ativação tangente hiperbólica.....	55
Figura 17 - Função de ativação Linear.....	56
Figura 18 - Rede Perceptron de Multicamadas .....	57
Figura 19 - Rede recorrente de Hopfield .....	59
Figura 20 - Máquina de Boltzmann .....	60
Figura 21 - Curva ABC aplicada a empresa escolhida.....	64
Figura 22 - Saída Gráfica do Método Média Móvel Simples .....	65
Figura 23 - Saída Gráfica do Método Média Móvel Ponderada.....	66
Figura 24 - Saída Gráfica do Método Amortecimento Exponencial Simples .....	66
Figura 25 - Saída Gráfica do Método Amortecimento Exponencial Dupla .....	67
Figura 26 - Saída Gráfica do Método Suavização Exponencial Com Tendência .....	68
Figura 27 - Saída Gráfica do Método Arima (1).....	68
Figura 28 - Saída Gráfica do Método Arima(2).....	69
Figura 29 - Saída Gráfica do Método Ma(1).....	70
Figura 30 - Saída Gráfica do Método Ma(2).....	70
Figura 31 - Saída Gráfica do Método ARMA(1,1).....	71

Figura 32 - Saída Gráfica do Método ARMA (1,2).....	72
Figura 33 - Saída Gráfica do Método ARMA(2,1).....	72
Figura 34 - Saída Gráfica do Método ARMA (2,2).....	73
Figura 35 - Saída Gráfica do Método Rede Neural com 1 camada oculta .....	74
Figura 36 - Saída Gráfica do Método Rede Neural com 2 camadas ocultas.....	75
Figura 37 - Tela com parametrização do algoritmo .....	84
Figura 38 - Tela do Software Stata com o algoritmo da Média Móvel.....	84

## LISTA DE TABELAS

Tabela 1 - Ranking dos Top 5 produtos que tem maior participação no faturamento da empresa .....	64
Tabela 2 - Parâmetros das Redes Neurais para Projeção do Produto L.....	74
Tabela 3 - Resultados obtidos através dos modelos de projeção de demanda .....	75

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	14
1.1 ÁREA DA ABEPRO.....	15
1.2 OBJETIVO GERAL.....	16
<b>1.2.1 Objetivo Específico</b> .....	16
1.3 METODOLOGIA.....	16
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	17
2.1 INDÚSTRIA 4.0.....	17
<b>2.1.1 Big Data</b> .....	20
<b>2.1.2 Internet das Coisas</b> .....	22
2.2 PREVISÃO DE DEMANDA .....	24
<b>2.2.1 Média Móvel Simples</b> .....	25
<b>2.2.2 Média Móvel Ponderada</b> .....	26
<b>2.2.3 Amortecimento Exponencial Simples</b> .....	27
<b>2.2.4 Suavização Exponencial Dupla (Método de Brown)</b> .....	27
<b>2.2.5 Suavização Exponencial com Tendência (Método de Holt)</b> .....	29
<b>2.2.6 Amortecimento Exponencial Triplo (Método Winter)</b> .....	30
<b>2.2.7 Modelo de Croston</b> .....	31
<b>2.2.8 Modelo de Box &amp; Jenkins</b> .....	32
<b>2.2.9 Modelos Estacionários</b> .....	33
<b>2.2.9.1 Modelos auto-regressivos</b> .....	33
<b>2.2.9.2 Modelos de médias móveis</b> .....	34
<b>2.2.9.3 Modelos auto-regressivo de médias móveis</b> .....	35
<b>2.2.10 Modelos Não Estacionários</b> .....	35
<b>2.2.10.1 Modelo ARIMA</b> .....	35
<b>2.2.10.2 Modelo SARIMA</b> .....	37
2.3 CURVA ABC.....	38
2.4 INTELIGÊNCIA ARTIFICIAL.....	41
<b>2.4.1 Redes Neurais</b> .....	45
<b>2.4.1.2 Rede Neural Artificial</b> .....	47
<b>2.4.1.2.1 Perceptron</b> .....	48
<b>2.4.1.2.1.2 Operações Lógicas e Regiões Linearmente Separáveis</b> .....	49
<b>2.4.1.2.2 Função de Ativação</b> .....	51
<b>2.4.1.2.2.1 Função de Etapa Binária</b> .....	52

2.4.1.2.2.2 Função Degrau Bipolar.....	53
2.4.1.2.2.3 Função Rampa Simétrica.....	54
2.4.1.2.2.4 Função Sigmoidal ou Logística.....	55
2.4.1.2.2.5 Função Tangente Hiperbólica.....	56
2.4.1.2.2.6 Função de Ativação Linear.....	56
2.4.1.2.3 Arquitetura de Redes Neurais Artificiais.....	57
2.4.1.2.3.1 Redes Neurais Feed-Forward.....	57
2.4.1.2.3.2 Redes Neurais Recorrentes.....	58
2.4.1.2.3.3 Redes Neurais Conectadas Simetricamente.....	59
2.4.1.2.3.3.1 Rede Hopfield.....	59
2.4.1.2.3.3.2 Máquinas de Boltzmann.....	60
2.4.1.2.4 Treinamento de Redes Neurais.....	62
2.4.1.2.4.1 Aprendizado de Redes Neurais.....	62
2.4.1.2.4.1.1 Aprendizado Supervisionado.....	63
2.4.1.2.4.1.2 Aprendizado Não Supervisionado.....	63
<b>3. TRABALHO DE CAMPO.....</b>	<b>63</b>
3.1 CURVA ABC.....	64
3.2 SAÍDAS DOS DIFERENTES MODELOS DE PROJEÇÃO DE DEMANDA.....	65
<b>3.2.1 Saída da Projeção Utilizado Média Móvel Simples.....</b>	<b>66</b>
<b>3.2.2 Saída da Projeção Utilizando Média Móvel Ponderada.....</b>	<b>66</b>
<b>3.2.3 Saída da Projeção Utilizando Amortecimento Exponencial Simples.....</b>	<b>67</b>
<b>3.2.4 Saída da Projeção Utilizando Amortecimento Exponencial Dupla.....</b>	<b>68</b>
<b>3.2.5 Saída da Projeção Utilizando Suavização Exponencial Com Tendência.....</b>	<b>68</b>
<b>3.2.6 Saída da Projeção Utilizando Método Ar(1).....</b>	<b>69</b>
<b>3.2.7 Saída da Projeção Utilizando Método Arima (2).....</b>	<b>70</b>
<b>3.2.8 Saída da Projeção Método Média Móvel 1 - Ma(1).....</b>	<b>70</b>
<b>3.2.9 Saída da Projeção Método Média Móvel 2 - Ma(2).....</b>	<b>71</b>
<b>3.2.10 Saída da Projeção Método ARMA(1,1).....</b>	<b>72</b>
<b>3.2.11 Saída da Projeção Método ARMA(1,2).....</b>	<b>72</b>
<b>3.2.12 Saída da Projeção Método ARMA(2,1).....</b>	<b>73</b>
<b>3.2.13 Saída da Projeção Método ARMA(2,2).....</b>	<b>74</b>
<b>3.2.14 Saída da Projeção Método Redes Neurais.....</b>	<b>74</b>
3.3 CONSOLIDAÇÃO DOS RESULTADOS OBTIDOS.....	76
<b>4. CONCLUSÃO.....</b>	<b>77</b>

<b>REFERÊNCIAS</b> .....	78
<b>ANEXO A – TELA DO SOFTWARE STATA</b> .....	85
<b>ANEXO B – SAÍDA DO SOFTWARE SSPS PARA PERCEPTRON COM UMA CAMADA OCULTA</b> .....	86
<b>ANEXO C – SAÍDA DO SOFTWARE SSPS PARA PERCEPTRON COM DUAS CAMADAS OCULTAS</b> .....	89

## 1. INTRODUÇÃO

O mercado sempre esteve sujeito a regras ditadas pelos consumidores, porém de formas diferentes. Antigamente, quando os produtos eram feitos por artesãos, os clientes possuíam uma relação estreita com os profissionais que fabricavam os itens demandados. Porém, foi com a primeira Revolução Industrial que surgiu a necessidade de padronização dos produtos e processos – o que acarretou o distanciamento entre fabricante e cliente. Segundo Martins e Laugeni (2015), o consumidor perdeu o poder de influenciar na definição dos produtos, sendo essa decisão responsabilidade dos projetistas que, em regra, não consultavam o mercado quanto às suas necessidades. Essa filosofia permeou as empresas até meados dos anos 1960, sendo chamada de *product-out*, isso é, a produção passou a ser em massa e a preocupação dos profissionais envolvidos no processo era gerar a maior produtividade possível – sendo de responsabilidade do setor de vendas colocar no mercado todo esse estoque gerado para os consumidores. No entanto, posteriormente a Segunda Guerra Mundial (1945) também foram desenvolvidas outras estratégias empresariais, como por exemplo o *market-in*, ou seja, as empresas passaram a dar importância cada vez maior à voz do mercado com o intuito de entregar ao cliente um produto de altíssima qualidade, que atendesse suas necessidades e que fosse produzido da maneira mais enxuta possível (reduzindo desperdícios de tempo, materiais, mão de obra, etc).

Atualmente, tendo em vista a competitividade do mercado, é imprescindível que empresas realizem planejamentos estratégicos, segundo Martins e Laugeni (2015), com o objetivo de determinar diretrizes de atuação nas áreas administrativas, financeira, de marketing e manufatura – visando vantagens competitivas. Os estoques, por exemplo, apesar de serem vistos como grandes vilões e responsáveis por altos custos das operações, são de extrema importância. Não se deve operar com estoque zero, é imprescindível que haja um bom planejamento para otimizar a produção e definir, de maneira mais assertiva, o tamanho de lote ideal para suprir as necessidades dos clientes e manter pouco capital imobilizado. De acordo com Monks (1997), a razão para a manutenção do estoque pode ser elencada em seis principais motivos:

- a) para atender os clientes com demandas variáveis (sazonais ou imediatas);
- b) proteger contra erros de suprimento, faltas e esgotamento de estoque;

- c) auxiliar no nivelamento das atividades de produção e melhorar as relações de trabalho (estabilizar o número de funcionários contratados);
- d) facilitar a produção de artigos diferenciados dentro das mesmas instalações;
- e) fornecer um meio de obter e manusear materiais em lotes econômicos e de ganhar descontos por quantidade;
- f) oferecer um meio de proteção contra as incertezas de entregas e preços futuros, como greves, aumento de preço e inflação.

Tendo em vista estes apontamentos, assim como a busca incessante por vantagens competitivas, é indubitável a relevância da análise e projeção de demanda dentro das organizações – representando, segundo Mancuzo (2003), o elo entre a disponibilidade de estoque e a necessidade de redução de custos. Projeção ou previsão é a arte de especificar informações significantes sobre o futuro e deve estar relacionada a um conjunto de métodos e ao conhecimento do previsor sobre o mercado e não puramente adivinhações (NARASIMHAN et al., 1995). A previsão realizada de forma adequada permite utilizar as máquinas corretamente, realizar a reposição de materiais no momento e quantidade correta e ajustar o processo produtivo para que todas as etapas ocorram em sincronia de forma harmônica. No entanto, as previsões estão sujeitas a erros em suas estimativas e o previsor deve ser cuidadoso tanto na coleta de dados quanto na escolha do modelo de previsão para minimizar esses erros. Por esse motivo, o presente trabalho busca realizar o estudo de diversos métodos de projeção de demanda com o objetivo de compará-los e definir qual o melhor modelo para a empresa selecionada.

### 1.1 ÁREA DA ASSOCIAÇÃO BRASILEIRA DE ENGENHARIA DE PRODUÇÃO

O presente trabalho de conclusão de curso se enquadra na sexta área da Associação Brasileira de Engenharia de Produção (ABEPRO), Pesquisa Operacional, e na sub-área de inteligência computacional. Esse segmento é voltado para a resolução de problemas reais envolvendo situações de tomada de decisão através de modelos matemáticos usualmente processados computacionalmente. Segundo a própria ABEPRO, “a pesquisa operacional procura introduzir elementos de objetividade e racionalidade nos processos de tomada de decisão, sem descuidar dos

elementos subjetivos e de enquadramento organizacional que caracterizam os problemas”.

## 1.2 OBJETIVO GERAL

Realizar um estudo da previsão de demanda através da aplicação dos principais modelos estatísticos e da aplicação de modelos de redes neurais que utilizam aprendizado de máquina, com base no histórico de vendas de um produto dos últimos 5 anos de uma empresa do setor de tecnologia - entendendo as diferenças de cada modelo e buscando minimizar o erro da previsão.

### 1.2.1 Objetivo Específico

Identificar o método de previsão com menor erro para poder realizar previsões mais assertivas na indústria de componentes eletrônicos. Desta forma, pode-se diminuir o estoque, reduzindo custos e possibilitando a aplicação deste capital em outros setores da empresa que possam melhor beneficiá-la.

## 1.3 METODOLOGIA

O método de pesquisa empregado para realizar a previsão de demanda de componentes eletrônicos para 2020 é de natureza quantitativa e de caráter exploratório – uma vez que o banco de dados será manuseado e será considerado apenas o histórico de vendas dos últimos 5 anos, desconsiderando aspectos socioeconômicos.

Os dados foram fornecidos por uma empresa de pequeno porte do setor de tecnologia e, para que os métodos estatísticos sejam aplicados corretamente, foi realizada uma pesquisa para identificar os principais e mais eficientes métodos de acordo com as características dos dados fornecidos. Os resultados, por outro lado, foram obtidos através de softwares como o Excel, Stata e SPSS e o melhor método selecionado com base no menor erro apresentado para a previsão do conjunto de dados com o objetivo de responder a seguinte pergunta de pesquisa: Qual o modelo de projeção de demanda que oferece resultados mais assertivos e que, conseqüentemente, auxiliam a otimizar o estoque da empresa do setor de tecnologia?

## 2. FUNDAMENTAÇÃO TEÓRICA

Essa seção tem como principal objetivo discutir o contexto de pesquisa no qual este trabalho se insere e enfatizar aspectos técnicos utilizados, como por exemplo as ferramentas de projeção de demanda.

### 2.1 INDÚSTRIA 4.0

Segundo Kang (2016), o termo Indústria 4.0 (I 4.0) foi introduzido pela primeira vez em 2011 na feira de Hanover, na Alemanha. O termo provém de um programa criado pelo governo alemão com o mesmo nome - que tinha como objetivo fomentar a automatização da manufatura e, conseqüentemente, aumentar a produtividade nas linhas de produção. Este incentivo a automatização acabou por gerar uma maior competitividade com as indústrias internacionais - através das fábricas inteligentes (*smart manufacturing*) e sendo parte integrante do movimento conhecido como a Quarta Revolução Industrial, que está inserida na era da informação digital. Uma parte fundamental dos processos industriais é a tecnologia da informação pois as decisões são tomadas automaticamente a partir do uso de um grande conjunto de dados armazenados, mais comumente conhecido como *Big Data* (FIRJAN, 2016).

De acordo com Veza (2015), a definição de indústria 4.0 engloba as políticas e práticas corporativas que levam a manufatura industrial em direção a Quarta Revolução Industrial. As Revoluções Industriais anteriores, através do uso de tecnologias de rápida disseminação, como mecanização, eletricidade e tecnologia da informação, alcançaram um nível de produtividade extrema.

A indústria 4.0 é considerada uma nova era industrial marcada pela capacidade de tomar decisões autônomas, pela utilização de sistemas inteligentes e pelo elevado grau de automação. Essa nova era industrial surge através do avanço tecnológico da internet e do desenvolvimento nos campos de objetos inteligentes derivados do crescente aumento da automação dos processos industriais. Os três elementos principais desta nova revolução industrial são: os sistemas cyber-físicos, o ciclo de vida do produto e a rede de produção e produto. O primeiro elemento através da integração dos sistemas MES (*Manufacturing Execution System*) e ERP (*Enterprise Resource Planning*) aumenta o nível de automação - possibilitando a total troca de informações. A partir de um modelo de dados uniforme é possível fazer a fusão do

ciclo de vida de produto e da produção explanando, assim, o segundo elemento. Esses dois elementos permitem que os requisitos resultantes de ciclos cada vez menores possam ser gerenciados técnica e economicamente (BRYNJOLFSSON, E.; MCAFEE, A., 2014).

A partir da utilização de sistemas cyber-físicos é possível integrar os mundos virtual e físico. Estes sistemas utilizam softwares integrados que realizam uma rápida troca de informação e, para realizar essa integração, é necessário que haja não só sensores e atuadores, mas como também um controle preciso do processo produtivo e que haja uma alta flexibilidade de processos. Sendo utilizados corretamente, estes sistemas possibilitam a resolução de diversos problemas da nossa sociedade – como por exemplo questões de segurança, mudanças climáticas, saúde, gestão de megacidades, entre muitos outros. De acordo com Hellinger e Seeger (2011), este avanço tecnológico é possibilitado pelo desenvolvimento de redes globais, como a internet, e pelo fácil acesso à informação através das nuvens de dados.

A integração é feita pela internet das coisas (IoT) que, de acordo com Hermann, Pentek e Otto (2016), é a conectividade de todas as partes do processo em tempo real - através da comunicação das máquinas de forma autônoma por meio de redes sem fio, sendo definidas tarefas e funções a serem desempenhadas para o correto funcionamento da fábrica.

Através da tecnologia RFID (*Radio Frequency Identification*) os produtos inteligentes conseguem deter as informações necessárias para o seu processo produtivo. Esta conectividade entre a fábrica inteligente e o produto inteligente é um dos destaques da I4.0. Segundo Hermann, Penntek e Otto (2016), a conexão funciona a partir de diversos sensores e atuadores conectados pela IoT - que enviam as informações para os meios de produção, possibilitando a autonomia de decisão, ou seja, a própria fábrica identifica e corrige problemas encontrados durante o processo produtivo. Diversos sensores, atuadores, máquinas e produtos comunicam-se por meio de wi-fi, estabelecendo uma troca de dados em tempo real entre todos os equipamentos envolvidos no processo.

De acordo com a *German Commission for Electrical, Electronic & Information Technologies of DIN and VDE (VDE-DKE)* - organização alemã responsável pelo desenvolvimento e adaptação de padrões de segurança nas áreas de engenharia elétrica e tecnologia de informação - o grande objetivo da I4.0 é aplicar os avanços adquiridos nas áreas de tecnologia da informação e nas áreas de comunicação no

ambiente das empresas industriais. Para Russwurn (2014), a utilização de sensores e chips de RFID possibilita que produtos, ferramentas e meios de produção troquem informações e, com o gerenciamento integrado de todas as partes da produção, é possível levar esta informação até mesmo para fora da fábrica.

O aumento de produtividade é possível através de vários facilitadores da I4.0, como por exemplo os mecanismos de colaboração. A mudança nas indústrias é possível devido ao alto potencial de crescimento da produtividade, porém a responsabilidade de tomar medidas lucrativas a partir de mudanças sociais e tecnológicas cai sobre a própria indústria produtora. É necessária a criação de pré-condições fundamentais no sistema de produção, divididas em dois níveis: a distinção entre componentes de *hardware* ou *software* e a alocação para o mundo virtual ou físico. Esta categorização leva a quatro pré-condições principais: globalização da tecnologia de informação, fonte única de verdade, automação e cooperação (SCHUH, G.,2015).

Estima-se que em 15 anos a indústria 4.0 deve movimentar US\$ 15 trilhões, mercado este que já está bem consolidado nos Estados Unidos da América e suas tendências são bem otimistas. De acordo com um relatório divulgado pela BP (um dos maiores grupos de energia do mundo), a previsão é que, até 2050, exista um potencial para aumentar os volumes de produção e reduzir custos em até 13% em alguns segmentos industriais. A Quarta Revolução Industrial está diretamente ligada com a coleta e a análise de grandes volumes de dados, mais comumente chamado de *Big Data*, e isto evidencia o fato de que um dos maiores princípios da I4.0 é a interoperabilidade, onde os fatores humanos e tecnológicos interagem com os dados assegurando, assim, a possibilidade da indústria ditar tendências enquanto é alimentada pelas necessidades dos consumidores. Não somente sendo tecnologia de ponta, mas também a alta interação e sinergia. Um dos maiores desafios de implementação destes sistemas é a necessidade de desenvolvimento de dispositivos inteligentes, pois as fábricas que operam de acordo com a I4.0 requerem uma elevada quantidade de dispositivos artificiais que reduzem o envolvimento de mão de obra no processo de produção - porém cada fábrica necessita de uma configuração diferente para realizar a sua produção. Para que isto seja resolvido, é necessário um grande investimento de tempo e de dinheiro antes mesmo da implementação destes dispositivos para que a fábrica opere nos padrões da I4.0. De acordo com Zhou (2015), todas as partes deste processo geram dados e a sua integração e análise

devem ser feitas de modo que todos os recursos da cadeia de produção sejam otimizados.

Segundo Stock (2016), a Quarta Revolução Industrial pode ser dividida em duas perspectivas: uma perspectiva macro que inclui a integração horizontal e a engenharia de ponta a ponta - que é caracterizada por uma *network* de módulos de criação, que por sua vez estão ligados em todo o processo de vida do produto, gerando assim uma *network* inteligente de módulos que engloba toda a cadeia de produção de diferentes produtos - criando uma nova categoria de modelos de negócios diferentes dos atuais.

Já a perspectiva micro inclui três integrações (horizontal, vertical e de ponta a ponta). Esta perspectiva é caracterizada pelas conexões de módulos de criação de valor ao longo do fluxo de produtos da fábrica inteligente, integrando uma logística inteligente que descreve o cruzamento dos fatores de criação de valor, equipamentos, pessoas e produtos. Estes módulos são sistemas cyber-físicos que incorporam equipamentos de produção que utilizam sistemas de sensores para identificar e localizar os fatores de criação de valor, monitorando os processos de fabricação. A partir da análise destes dados os atuadores aplicados ao sistema podem reagir em tempo real a mudanças específicas do produto. Segundo Stock (2016), a comunicação e troca de dados inteligentes entre os fatores de criação de valor, o módulo de criação de valor e o equipamento de transporte, são executados através de sistemas em nuvem.

Em suma, a estratégia da I4.0 se baseia em manufaturas inteligentes utilizando sistemas cyber-físicos para mudar de uma produção centralizada para uma produção descentralizada, sendo flexível e complexa e envolvendo tecnologias nas áreas de manufatura digital, comunicação em rede, automação e outras. De acordo com Zhou (2015), a sua implementação é baseada em design digital, simulações, processos altamente automatizados, gerenciamento dos dados da produção em rede e no processo da produção.

### **2.1.1 Big Data**

*Big Data* pode ser entendido como um termo que se refere a um conjunto de dados que é tão grande e complexo que é difícil ou até mesmo impossível de processá-lo através de métodos tradicionais. O volume, no entanto, não é o que torna

o *Big Data* uma ferramenta essencial dentro das organizações e sim a maneira como a empresa lida e faz com esse conjunto de informações sejam geradas – que pode auxiliar a organização na tomada de decisões e planejamentos estratégicos.

O ato de acessar e estocar um grande número de informações para análise não é algo recente, no entanto, o conceito de *Big Data* ganhou relevância em 2001 quando o analista Doug Laney articulou o conceito que conhecemos hoje como “Os Três V’s”: Volume, Velocidade e Variedade. O primeiro se refere à enorme quantidade de dados provenientes de diversas fontes; O segundo é a constatação de que os dados fluem em uma velocidade nunca antes vista e que devem ser processados em tempo hábil; Já o terceiro, se refere a gama de dados existentes – que podem ser tanto dados estruturados como até mesmo documentos de texto, áudio, vídeo, transações financeiras, etc (dados não estruturados). Esses dados podem ser categorizados em três principais tipos, segundo Nascimento (2017), sendo eles: *Social Data*, *Enterprise Data* e *Data Of Things*. Os dados do tipo *Social Data* provém das pessoas, isso é, dos comportamentos, sejam eles no âmbito virtual ou não, que geram informações úteis para as organizações. Com esses dados é possível entender a maneira como os internautas, consumidores e potenciais clientes se comportam, que tipo de informação procuram, qual o padrão de compra, etc para que sua experiência online e física seja cada vez mais personalizada. Já o *Enterprise Data* são os dados gerados pelas empresas, como dados do financeiro, operações, recursos humanos, etc. Apesar de algumas organizações negligenciarem esse tipo de informação, elas são extremamente importantes para avaliar a produtividade dos setores, descobrir possíveis gargalos de operações e até mesmo propor melhorias e planejar os passos a serem seguidos pela empresa. Por outro lado, o *Data Of Things*, que será abordado no sub-tópico a seguir, é toda a informação gerada por dispositivos que estão conectados à internet e interagem entre si, como por exemplo carros, TV’s e geladeiras. Através dessa troca de informações em tempo real, é possível gerar novos dados capazes de auxiliar, por exemplo, no dia a dia dos motoristas no trânsito, como no caso do Waze ou Google Maps. Em suma, o *Big Data* é uma ferramenta potente, capaz de gerar um grande diferencial competitivo para aqueles que souberem utilizar o banco de dados corretamente. Para isso, é crucial tratar e organizar os dados para que sejam geradas informações úteis – que contribuam para a tomada de decisões.

### 2.1.2 Internet das Coisas

Segundo Xia (2012), estamos testemunhando o início de uma nova era da Internet, uma extensão que possibilita que os objetos comuns em nosso cotidiano se conectem à internet – permitindo que sejam acessados de maneira remota. Essa nova era é chamada de Internet das Coisas (*Internet of Things*, IoT) e é baseada na ideia de ubiquidade, ou seja, algo que está presente em todos os lugares e em todos os momentos, persistente, sempre disponível e atuante.

Devido ao rápido avanço das tecnologias subjacentes, a IoT está possibilitando enormes oportunidades para um grande número de novas aplicações que prometem melhorar a qualidade de nossas vidas. Nos últimos anos, a IoT ganhou muita atenção de pesquisadores e profissionais de todo o mundo. Segundo uma pesquisa realizada pela PwC em 2016, espera-se que até 2020, 72% das organizações no Brasil estejam digitalizadas. Os dados mostram que o fenômeno do IoT é uma realidade mundial, que afetará não somente as organizações, onde as máquinas serão mais autônomas e se comunicarão entre si, mas também os postos de trabalho tradicionais.

Apesar dos diversos impactos que a IoT tem e terá sobre a sociedade, é válido ressaltar que o conceito de Internet das Coisas é bastante recente – sendo empregado pela primeira vez em 1999, por Kevin Ashton. O britânico realizou uma apresentação para a renomada Procter & Gamble (P&G) integrando o uso de tecnologias de endereçamento de dados e sinais com a internet (COSTA, 2018). Ashton tinha como objetivo melhorar o fluxo de produtos e informações sem que houvesse a interferência humana direta e, para isso, utilizou a tecnologia de Identificação de Rádio Frequência (RFI) – onde foi possível constatar um ganho significativo em cadeias logísticas, permitindo movimentações confiáveis, rápidas e sincronizadas com todos os agentes envolvidos ao longo da cadeia por meio do compartilhamento de informações. Atualmente sabe-se que a internet é totalmente dependente do ser humano para obter dados, mas é também uma ferramenta facilitadora - gravando e recuperando quantias significativas de dados do mundo físico (ASHTON, 2009).

Para que a IoT seja viabilizada, segundo Costa (2018), é necessária uma combinação de diversas tecnologias: Computação, Serviços, Semântica, Identificação, Atuadores e Comunicação. Essas tecnologias podem ser descritas como a seguir, de acordo com Costa (2018):

- a) Computação: Engloba o sistema operacional de processamento, ou seja, dispositivos que executam os algoritmos locais nos objetos inteligentes;
- b) Serviços: Para que um serviço seja entregue, é necessário que haja identificação (mapeia entidades físicas de interesse do usuário em entidades virtuais), agregação de dados (precisa ser versátil para lidar com dados homogêneos ou não captados dos objetos inteligentes), colaboração inteligente (se refere a reação a um determinado cenário imposto) e ubiquidade (colaboração inteligente em qualquer lugar ou momento);
- c) Semântica: Pode ser entendida como a competência de extrair as informações dos objetos conectados – determinando os serviços com base nos dados coletados;
- d) Identificação: É o que informa se o objeto é inteligente ou não, isso é, se tem a capacidade de se conectar a rede ou não. As tecnologias mais usuais são RFID, *Near Field Communication* (NFC) e endereçamento IP;
- e) Atuadores: Após a coleta e armazenamento das informações pelos objetos inteligentes, os dados são encaminhados para *data warehouse*, nuvens ou centros de armazenamento. Com isso, os atuadores conseguem manipular o ambiente e reagir de acordo com as informações para ele fornecidas;
- f) Comunicação: É a tecnologia empregada para conectar os objetos inteligentes e é considerado um fator crítico por estar diretamente ligado ao consumo de energia dos aparelhos. As tecnologias mais usuais são o WiFi, Bluetooth e RFID.

Todas as pesquisas, as quais apontam para um potencial e rápido crescimento da Internet das Coisas, permitem inferir que haverá oportunidades únicas para os fabricantes tradicionais de equipamentos e eletrodomésticos transformarem seus produtos em “coisas inteligentes”. A IoT e serviços relacionados globalmente necessitam de Serviço de Internet Provedores (ISPs) para provisionar suas redes para fornecer qualidade de serviço para uma mistura de M2M (máquina para máquina), pessoa para máquina (P2M) e pessoa para pessoa (P2P).

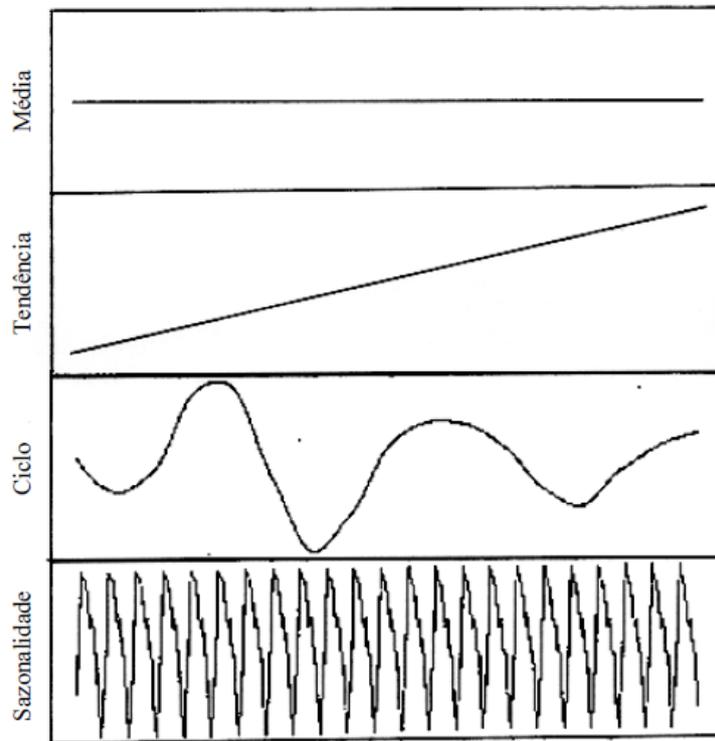
## 2.2 PREVISÃO DE DEMANDA

Realizar a previsão de demandas futuras tornou-se uma atividade essencial dentro das organizações tendo em vista o mercado cada vez mais competitivo. De acordo com Martins e Laugeni (2015), a previsão de demanda pode ser entendida como um processo metodológico para a determinação de dados futuros baseados em modelos estatísticos, econométricos ou até mesmo modelos subjetivos apoiados em uma metodologia de trabalho clara e previamente definida. A previsão de demanda tem como finalidade auxiliar na estruturação das organizações com o objetivo de dimensionar a produção de bens ou serviços de forma a atender a demanda dos consumidores. Segundo Makridakis (1988), essa atividade é parte integrante do processo decisório da gestão empresarial e, por permitir o planejamento de recursos, torna-se uma atividade base para o planejamento estratégico da produção, vendas e finanças.

No âmbito do Planejamento e Controle da Produção (PCP), as previsões costumam ser classificadas de acordo com o horizonte de planejamento, que podem ser de longo, médio ou curto prazo. De acordo com Fernandes e Godinho (2017), na perspectiva do PCP, as visões de longo prazo são importantes para o planejamento de novas instalações, produtos e gastos de capital. Já as visões de médio prazo servem como base para o planejamento agregado da produção e análises de capacidade agregadas. Por último, as visões de curto prazo auxiliam na programação da força de trabalho, de compras, nas análises de capacidade de curto prazo, entre outros.

As projeções são realizadas através de modelos matemáticos – como pode ser visto na subseção a seguir - e a escolha adequada, conforme as características da série de dados, é fundamental para que haja uma projeção mais assertiva. O emprego de cada modelo depende do comportamento da série temporal – a qual pode exibir até quatro características diferentes em seu comportamento, segundo Makridakis (1988): média, sazonalidade, ciclo e tendência – como pode ser visto na Figura 1 abaixo:

Figura 1 - Comportamento de demandas



Fonte: PELLEGRINI, F. R. (2000)

### 2.2.1 Média Móvel Simples

No método da média móvel simples, a projeção no período futuro  $t$ , de acordo com Martins e Laugeni (2015), é calculada como sendo a média de  $n$  períodos anteriores. Utiliza-se valores determinados, geralmente os mais recentes, e a cada período de atualização da previsão, os valores mais antigos são substituídos. Desta forma, temos que:

$$\hat{y}_t = \frac{\bar{y}_{t-1} + \bar{y}_{t-2} + \bar{y}_{t-3} + \dots + \bar{y}_{t-n}}{n} \quad (1)$$

Sendo:

$\hat{y}$  : valores projetados

$\bar{y}$  : valores realizados

$t$  : período

$n$  : número de períodos

Segundo Corrêa et.al (2011), o número de períodos  $t$  exerce grande influência na suavização das variações aleatórias e sensibilidade do modelo a mudanças no comportamento de venda. Isto é, quanto maior o número de períodos, maior a suavização das variações aleatórias e menor a sensibilidade do modelo. Ainda de acordo com os mesmos autores, o método de média móvel simples é recomendado nos casos quando se busca um modelo de projeção de vendas simples e de baixo custo para prever a demanda de itens com histórico de pequenas flutuações e sem indícios de tendências. Isso se deve ao fato de que, em séries históricas que apresentam tendência ou sazonalidade, é necessário avaliar um maior número de dados e não desconsiderar os valores anteriores para adição de novos, como no caso da média simples. Uma alternativa para suavizar esse erro, segundo Davis, Aquilano e Chase (2001), é a utilização de outro método de projeção, como a média ponderada – porém é de extrema importância que haja um conhecimento sobre os dados a serem manipulados para que sejam determinados os pesos mais adequados, como é possível verificar no próximo tópico.

### 2.2.2 Média Móvel Ponderada

Nesse método, diferentemente da média móvel simples, atribui-se um peso diferente a cada um dos dados – sendo que a soma desses pesos deve ser igual a 1 e os maiores devem ser atribuídos às vendas mais recentes.

$$\hat{y}_t = \frac{\bar{y}_{t-1} \times p_1 + \bar{y}_{t-2} \times p_2 + \bar{y}_{t-3} \times p_3 + \dots + \bar{y}_{t-n} \times p_n}{p_1 + p_2 + \dots + p_n} \quad (2)$$

Onde:

$p_1, p_2 \dots p_n$  : são os pesos relativos

$\hat{y}$  : valores projetados

$\bar{y}$  : valores realizados

$t$  : período

$n$  : número de períodos

### 2.2.3 Amortecimento Exponencial Simples

Assim como no modelo de Média Móvel Ponderada, o grau de importância de cada observação decresce com o tempo – mas os modelos se diferem pelo fato de que nesse método os pesos decrescem exponencialmente do tempo presente em direção ao passado. Segundo Tubino (2000), “em sua forma de apresentação mais simples, cada nova previsão é obtida com base na previsão anterior, acrescida do erro cometido corrigido por um coeficiente de ponderação”. Desta forma, temos que:

$$P_{t+1} = \beta R_t + \beta(1 - \beta)P_t \quad (3)$$

Sendo:

$P_{t-1}$  : previsão para o período seguinte

$R_t$  : valor real observado no período  $t$

$\beta$  : coeficiente de amortecimento ( $0 \leq \beta \leq 1$ )

$P_t$  : previsão referente ao período  $t$

O coeficiente de ponderação  $\beta$ , que pode variar entre 0 e 1, revela a rapidez com a qual o modelo reage a uma variação real da demanda, isto é, quanto maior o valor de  $\beta$ , mais rápido o modelo reagirá. No entanto, é necessário estabelecer um valor adequado para este coeficiente - uma vez que valores muito altos sujeitam as previsões a variações aleatórias da demanda e valores muito pequenos podem fazer com que as previsões fiquem discrepantes da demanda real. Por conta disso, é usual, segundo Pereira et.al (2006), que pacotes computacionais para gerenciamento e previsão de demanda incluam simulações para ajustar o valor de  $\beta$ . Pelo fato de armazenar apenas 3 informações por item, isto é, previsão passada, demanda e coeficiente de ponderação  $\beta$ , o modelo de média exponencial móvel é considerado um dos mais comuns e simples de serem utilizados por sistemas computacionais de acordo com Tubino (2000).

### 2.2.4 Suavização Exponencial Dupla (Método de Brown)

Em casos em que há tendência, o modelo de suavização exponencial dupla (Método Brown) pode ser utilizado para as previsões de séries temporais. Esse

modelo fornece um método que ao mesmo tempo leva em consideração o tempo (maior peso para dados mais recentes) e a tendência. De acordo com Guimarães (2008), o primeiro passo é calcular as suavizações:

$$S_0 = a_0 - \left(\frac{1 - \alpha}{\alpha}\right) b_0 \quad (4)$$

$$\hat{S}_0 = a_0 - 2\left(\frac{1 - \alpha}{\alpha}\right) b_0 \quad (5)$$

Sendo:

$S_0$  : primeira suavização para o período 0

$\hat{S}_0$  : segunda suavização para o período 0

$a_0$  : coeficiente linear da regressão dos valores da série pelos números dos períodos

$b_0$  : coeficiente angular da regressão dos valores da série pelos números dos períodos

$\alpha$  : coeficiente de suavização da média

Com isso, é possível calcular as suavizações para o período t:

$$S_t = \alpha D_t + (1 - \alpha) S_{t-1} \quad (6)$$

$$\hat{S}_t = \alpha S_t + (1 - \alpha) \hat{S}_{t-1} \quad (7)$$

Sendo:

$D_t$  = demanda do período t

O fator de ajuste adicional, portanto, pode ser calculado através das diferenças entre as duas suavizações:

$$\alpha_t = 2S_t - \hat{S}_t \quad (8)$$

$$b_t = \left(\frac{\alpha}{1 - \alpha}\right) (S_t - \hat{S}_t) \quad (9)$$

Por fim, a previsão de demanda para os próximos  $p$  períodos pode ser determinada por:

$$P_{t+p} = a_t + b_t p \quad (10)$$

### 2.2.5 Suavização Exponencial com Tendência (Método de Holt)

Esse modelo, assim como o modelo de suavização dupla, é indicado para séries temporais que apresentam tendência, porém com a diferença de que o modelo de suavização exponencial com tendência (Método de Holt) é adequado para séries temporais sem sazonalidade. De acordo com Guimarães (2008), o primeiro passo é calcular a média e a tendência relativa dos dados da série temporal para que então seja estimada a primeira previsão. A tendência, no entanto, é gradual e em longo prazo e por isso é plausível supor um comportamento linear entre a demanda e o tempo. Dessa forma, temos:

$$D_t = a + b_t \quad (11)$$

Onde:

$D_t$  : demanda no período  $t$

$a$  : estimativa inicial da média (mede a estimativa da demanda no período  $t=0$ )

$b$  : estimativa inicial da tendência (mede a taxa de inclinação 0 da demanda em função do período  $t$ )

Desta forma, a previsão de demanda para o período seguinte pode ser calculada através da soma da média e tendência relativas ao período anterior com os ajustes exponenciais:

$$P_{t+1} = N_t + T_t \quad (12)$$

$$N_t = P_t + \alpha(D_t - P_t) \quad (13)$$

$$T_t = T_{t-1} + \beta((P_t - P_{t-1}) - T_{t-1}) \quad (14)$$

Sendo:

$P_{t+1}$  : estimativa de previsão de demanda para o período t+1

$P_t$  : estimativa de previsão de demanda para o período t

$P_{t-1}$  : estimativa de demanda para o período t-1

$N_t$  : estimativa da média para o período t

$T_t$  : estimativa da tendência exponencial para o período t

$T_{t-1}$  : estimativa da tendência exponencial para o período t-1

$D_t$  : demanda no período t

$\alpha$  : coeficiente de suavização da média ( $0 \leq \alpha \leq 1$ )

$\beta$  : coeficiente de suavização da tendência ( $0 \leq \beta \leq 1$ )

Os coeficientes de variação, como mencionado anteriormente, expressam o grau de reação do modelo, sendo os valores próximos de 0 para um modelo menos reativo e os valores mais próximos a 1 para modelos mais reativos.

### 2.2.6 Amortecimento Exponencial Triplo (Método Winter)

Segundo Martins (2011), o método de Winter é um método de amortecimento exponencial que leva em conta os componentes de sazonalidade da série de dados observados. Por isto, é adequado para previsão de séries que apresentam tendências e sazonalidades e pode ser expressado da seguinte maneira:

$$s_t = \gamma \left( \frac{R_t}{N_t} \right) + (1 + \gamma) s_{t-c} \quad (15)$$

$s_t$  : Componente Sazonal

$\left( \frac{R_t}{N_t} \right)$  : Ajuste Sazonal calculado no período t

$s_{t-c}$  : Ajuste sazonal calculado c períodos atrás. Para previsão mensal (semanal) e sazonalidade ao longo do ano (mês), usa-se  $c = 12$  (4).

$\gamma$  : Coeficiente de amortecimento para a estimativa da sazonalidade ( $0 \leq \gamma \leq 1$ )

Depois, calcula-se a tendência (Holt):

$$T_t = \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1} \quad (16)$$

E o nível considerando o ajuste sazonal:

$$N_t = \alpha \left( \frac{R_t}{S_{t-c}} \right) + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (17)$$

Agora pode-se, por fim, realizar o cálculo da previsão:

$$P_{t+p} = (N_t + pT_t)S_{t-c+p} \quad (18)$$

$T_t$  : Componente Tendência

$\alpha$  : Coeficiente de Amortecimento

$\beta$  : Coeficiente de amortecimento para a estimativa de tendência

$R_t$  : Valor real observado no período t

p : Número de períodos a serem projetados

### 2.2.7 Modelo de Croston

De acordo com Koenig (2014), o método de Croston executa, separadamente, duas suavizações exponenciais. Primeiramente, utiliza a suavização exponencial simples sobre a demanda - estimando a demanda média futura. Em seguida, aplica-se a suavização exponencial simples sobre o intervalo entre as demandas, estimando o intervalo médio futuro. O método atualiza as estimativas após a ocorrência de demanda. Se não ocorrer demanda no espaço entre uma revisão do período e outra (período t), o método somente incrementa a contagem dos períodos desde a última demanda. Para que fosse possível esta contagem dos períodos entre as demandas, Croston introduziu a variável q, que será explicada a seguir. Têm-se então duas situações possíveis:

a) Para períodos sem demanda ( $y_t = 0$ )

$$p_t = p_{t-1} \quad z_t = z_{t-1} \quad q = q + 1 \quad (19)$$

b) Para períodos com demanda ( $y_t \neq 0$ )

$$p_t = p_{t-1} + \alpha(q - p_{t-1}) \quad (20)$$

$$z_t = z_{t-1} + \alpha(y - z_{t-1}) \quad (21)$$

$$q = 1 \quad (22)$$

Combinando a previsão do tamanho da demanda com a previsão do intervalo entre demandas, obtém-se a equação central proposta por Croston:

$$y_t^* = \frac{z_t}{p_t} \quad (23)$$

$y_t$  : Demanda de um item no tempo t

$q$  : Intervalo de tempo desde a última demanda

$p_t$  : Estimativa de Croston de intervalo médio entre transações

$\alpha$  : Parâmetro de suavização entre ( $0 \leq \alpha \leq 1$ )

$z_t$  : Estimativa de Croston do tamanho médio da demanda

$y_t^*$  : Previsão da Demanda

### 2.2.8 Modelo de Box & Jenkins

O método de Box & Jenkins tem recebido muita atenção, segundo Morettin e Toloí (2006), por ser utilizado na análise de modelos paramétricos. Esse método consiste em ajustar modelos auto-regressivos integrados de médias móveis, ou simplesmente ARIMA (p, d, q), a um conjunto de dados. A estratégia de construção

do modelo deve ser baseada em um ciclo iterativo, onde a estrutura será ter a como base os próprios dados. Os passos do ciclo iterativo são:

- a) Considerar uma classe geral de modelos para análise (especificação);
- b) Identificação de um modelo, com base na análise de auto correlações, auto correlações parciais, dentre outros critérios;
- c) Estimação dos parâmetros do modelo identificado;
- d) Verificação do modelo ajustado, por meio de análise de resíduos, para verificar se o modelo está adequado para os propósitos estabelecidos.

Caso o modelo não seja o adequado, será necessário repetir esse ciclo, começando pela identificação. Muitas vezes identifica-se mais de um modelo a ser estimado e verificado, e dentre eles deve-se escolher o que melhor se encaixa. A dificuldade para aplicar esse modelo é que requer a utilização de programas adequados. (MORRETIN; TOLOI, 2006).

Os modelos Box & Jenkins, de acordo com Bacci (2007), se dividem em: Modelos de Médias Móveis (MA), Modelos Auto-Regressivos (AR), Modelo Misto Auto-Regressivo e de Médias Móveis (ARMA) e Modelo Misto Auto-Regressivo e de Médias Móveis para processos não estacionários homogêneos (ARIMA).

### **2.2.9 Modelos Estacionários**

Segundo Werner e Ribeiro (2003), os modelos estacionários são aqueles que pressupõem que o modelo se encontra em “equilíbrio”. É considerado fracamente estacionário se suas média e variância se mantêm inalteradas com o passar do tempo e a função auto covariância depende somente da defasagem entre instantes de tempo. É considerado fortemente estacionário se todos os conjuntos se mantêm inalterados com o passar do tempo.

#### **2.2.9.1 Modelos auto-regressivos**

De acordo com Pindyck e Rubinfeld (2013) no processo de auto-regressão de ordem  $p$  a atual observação  $y_t$  é gerada pela média ponderada das observações passadas voltando  $p$  períodos, juntamente com uma distribuição aleatória no período

atual. Esse processo é denominado AR( $p$ ) e sua equação pode ser escrita da seguinte forma:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \delta + \varepsilon_t \quad (24)$$

Onde:

$\delta$  : termo constante que se relaciona para a média do processo estocástico.

Uma das condições necessárias para que o modelo AR( $p$ ) seja estacionário é que a somatória dos  $\Phi$  tenha um valor menor que 1, com  $\Phi$  invariáveis em função do tempo. Porém, essa condição só é suficiente para afirmar que o modelo seja estacionário caso os  $\Phi$  sejam finitos, caso contrário é necessário outra condição para validar o modelo.

### **2.2.9.2 Modelos de médias móveis**

Segundo Pindyck e Rubinfeld (1991), no processo de médias móveis de ordem  $q$ , cada observação  $y_t$  é gerada pela média ponderada das perturbações aleatórias voltando  $q$  períodos. Esse processo é denominado MA( $q$ ) e sua equação pode ser escrita da seguinte forma:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (25)$$

Onde os parâmetros  $\theta_1, \theta_2, \dots, \theta_q$  podem ser tanto positivos quanto negativos e determinam os pesos da média ponderada.

Nesse modelo cada termo de perturbação aleatória  $\varepsilon_t$ , gerado pelo ruído branco, é considerado uma variável normal aleatória – o que significa variância e covariância iguais a zero.

Pindyck e Rubinfeld (2013) demonstram que o processo de médias móveis de período definido como o MA(2), de dois períodos, tem memória de exatamente dois períodos, portanto o valor de  $y_t$  é influenciado somente pelo evento do período atual, um período antes e dois períodos antes.

### 2.2.9.3 Modelos auto-regressivo de médias móveis

Segundo Morettin e Tolo (2006) e Pindyck e Rubinfeld (2013), o modelo auto-regressivo de médias móveis é adequado para muitas séries reais pelo fato de terem a inclusão de termos auto-regressivos e de médias móveis, partindo de uma ordem  $(p, q)$ . Esse processo pode ser escrito pela equação a seguir:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (26)$$

Assume-se que o modelo ARMA( $p, q$ ) seja estacionário quando a somatória dos  $\phi$  tenha um valor menor que 1 e seja invariável com o tempo.

### 2.2.10 Modelos Não Estacionários

Segundo Morettin e Tolo (2006), as séries que podem ser encontradas no dia-a-dia não são estacionárias, com exceção de séries subjacentes de fenômenos físicos, como os ruídos. Quando uma série apresenta média e variância dependentes do tempo é possível classificá-las como não estacionárias, de acordo com Werner e Ribeiro (2003). Essa não-estacionariedade impacta a série de duas formas:

- a) existe uma inclinação nos dados e não permanecem ao redor de uma linha horizontal ao longo do tempo e/ou
- b) a variação dos dados não permanece constante sobre tempo, ou seja, as flutuações aumentam e diminuem com o passar do tempo, que indica que a variância está se alterando.

Séries econômicas são não-estacionárias, porém quando diferenciadas entre si são estacionárias (MORETTIN e TOLOI, 2006).

#### 2.2.10.1 Modelo ARIMA

O modelo ARIMA, de acordo com Brownlee (2017), é pertencente a classe de modelos estatísticos de análise e previsão de série de dados temporais e, mesmo sendo um modelo simples, atende as estruturas padrões e tem uma previsão hábil. A sigla desse modelo são os aspectos chaves do mesmo, sendo eles:

- a) AR: Autorregressão. Onde se utiliza a relação de dependência entre observações e observações atrasadas.
- b) I: Integrado. O uso da diferenciação de uma observação inicial para tornar o modelo estacionário.
- c) MA: Média Móvel. A utilização da dependência entre observação e erro residual de uma média móvel aplicada a observações atrasadas.

Ainda segundo Brownlee (2017), cada componente desses é um parâmetro do modelo, que é a notação padrão e é usado para ARIMA ( $p, d, q$ ) - onde cada parâmetro é substituído por um valor inteiro para indicar rapidamente o modelo que está sendo usado. A letra  $p$  refere-se ao número de observações atrasadas; a letra  $d$  refere-se à quantidade de diferenciações feitas de uma observação inicial; e a letra  $q$  refere-se ao tamanho da média móvel.

O modelo é construído com base no número e tipo de termo especificado e os dados são preparados de acordo com o grau de diferenciação no intuito de ser estacionário, ou seja, remover tendência e sazonalidades que afetam o modelo negativamente. Quando um dos parâmetros tem seu valor igual a zero, significa que não será utilizado - transformando assim em outro modelo mais simples como ARMA, AR, I ou MA.

Segundo Pindyck e Rubinfeld (2013), a série  $y_t$  é um processo integrado de auto-regressão de média móvel ( $p, d, q$ ), ou simplesmente ARIMA( $p, d, q$ ) e pode ser descrito como:

$$\phi(B)\Delta^d y_t = \delta + \theta(B)\varepsilon_t \quad (27)$$

Sendo:

$$w_t = \Delta^d y_t \quad (28)$$

Onde  $d$  é a ordem da série  $W_t$  e representa o total de vezes que a série  $y_t$  foi diferenciada até se transformar em uma série estacionária,  $\phi(B)$  é o operador de auto-regressão e  $\theta(B)$  é o operador de média móvel.

### 2.2.10.2 Modelo SARIMA

O modelo SARIMA, segundo Werner e Ribeiro (2003), apresentam auto-correlação sazonal, sendo assim, contém tanto parte não sazonal ( $p, d, q$ ) como também sazonal ( $P, D, Q$ ) e sua equação geral pode ser escrita como:

$$\begin{aligned} (1 - \Phi_1 L - \dots - \Phi_p L^p)(1 - \Phi_1 L^s - \dots - \Phi_p L^{ps})(1 - L)^d(1 - L)^D Z_t \\ = (1 - \theta_1 L - \dots - \theta_q L^q)(1 - \theta_1 L^s - \dots - \theta_q L^{qs}) \varepsilon_t \end{aligned} \quad (29)$$

Onde:

$$(1 - \Phi_1 L - \dots - \Phi_p L^p) \quad (30)$$

é a parte auto regressiva não sazonal de ordem  $p$ ;

$$(1 - \Phi_1 L^s - \dots - \Phi_p L^{ps}) \quad (31)$$

é a parte auto regressiva sazonal de ordem  $P$  e estação sazonal  $s$ ;

$$(1 - L)^d \quad (32)$$

é a parte de integração não sazonal de ordem  $d$ ;

$$(1 - L^s)^D \quad (33)$$

é a parte de integração sazonal de ordem  $D$  e estação sazonal  $s$ ;

$$(1 - \theta_1 L - \dots - \theta_q L^q) \quad (34)$$

é a parte não sazonal de médias móveis de ordem  $q$ ;

$$(1 - \theta_1 L^s - \dots - \theta_q L^{qs}) \varepsilon_t \quad (35)$$

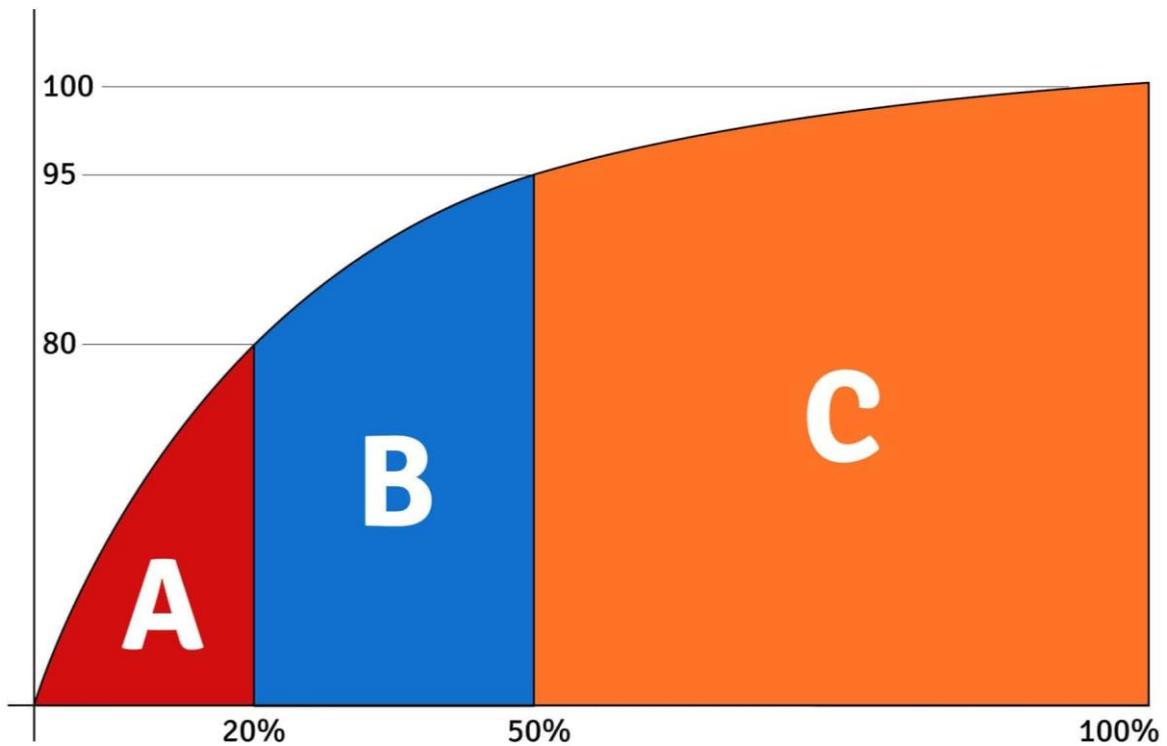
é a parte sazonal de médias móveis de ordem  $Q$  e estação sazonal  $s$ .

### 2.3 CURVA ABC

A Curva ABC é fundamentada pelos princípios de Pareto – economista italiano, do século XIX, que desenvolveu a teoria de que o padrão de distribuição de renda é histórica e geograficamente constante, não dependendo de políticas tributárias e de bem-estar social. Através de seus estudos Vilfredo Pareto observou que uma pequena parcela da população, cerca de 20%, detinha a maior parte da riqueza, isso é, mais de 80%. Suas conclusões foram demonstradas graficamente, em uma curva acumulativa, que posteriormente ficou conhecida como Curva ABC, ou também por Curva 80-20. De acordo com Mitiuye, Silva e Pereira (2008), a teoria do economista não se limitou apenas ao campo da economia e foi adaptada por muitos estudiosos, como por exemplo os estatísticos para áreas como a produção e serviços.

No setor de produção, por exemplo, os princípios de Pareto possibilitam analisar o portfólio de uma organização a partir de seu faturamento e vice-versa para entender de que maneira a empresa pode estar vulnerável caso haja dependência de determinados itens. Isso é, os princípios nos permitem inferir que 80% dos produtos seriam responsáveis por 20% do faturamento, ao mesmo tempo que 20% do total de produtos seriam responsáveis por 80% do faturamento. Ainda segundo o ponto de vista da gestão de produtos, segundo Moreira (2008) a curva ABC também possibilita o controle de itens estocados a partir do critério de investimento utilizado em cada um. Já para Braga (1989) a curva ABC é mais bem utilizada quando o estoque é constituído por milhares de itens diferentes onde o grau de importância de alguns é maior para a empresa. Por fim, para Dias (2005) a curva possibilita a identificação dos produtos que requerem uma atenção maior da empresa, ranqueando prioridades e estabelecendo uma melhor política de vendas.

Figura 2 - Curva ABC



Fonte: MARMELO, D. (2019)

De acordo com Carvalho (2002), os três componentes da curva ABC que possibilitam não só realizar as análises, mas como também plotar graficamente, podem ser divididos da seguinte maneira: Na Classe A seriam os itens com maior grau de importância, valor ou quantidade e que devem receber atenção especial da empresa – correspondendo a 20% do total; Os itens representados na Classe B seriam os itens intermediários entre a classe A e a C, isso é, possuem um valor intermediário no faturamento da empresa – representando cerca de 30% do total; Já na Classe C estariam a maior parte dos itens, mas que possuem menor valor investido, ou seja, os itens que possuem menor impacto no faturamento da empresa – representando cerca de 50% do total

É evidente que não só as análises mencionadas, mas como também muitas outras possíveis, devem ser realizadas, se possível em planilhas eletrônicas – uma vez que essas são capazes de tornarem as análises muito mais dinâmicas. Através do cruzamento de informações como por exemplo o faturamento por produto e a margem de lucro, é possível plotar o gráfico da curva ABC tanto de produtos quanto de clientes. De acordo com Cobra (2005), esse cruzamento de informações de

clientes, por exemplo, possibilita determinar os produtos comprados por cada cliente permitindo chegar a diagnósticos relevantes tanto sobre os produtos quanto sobre os clientes que representam "o ganha pão" da empresa.

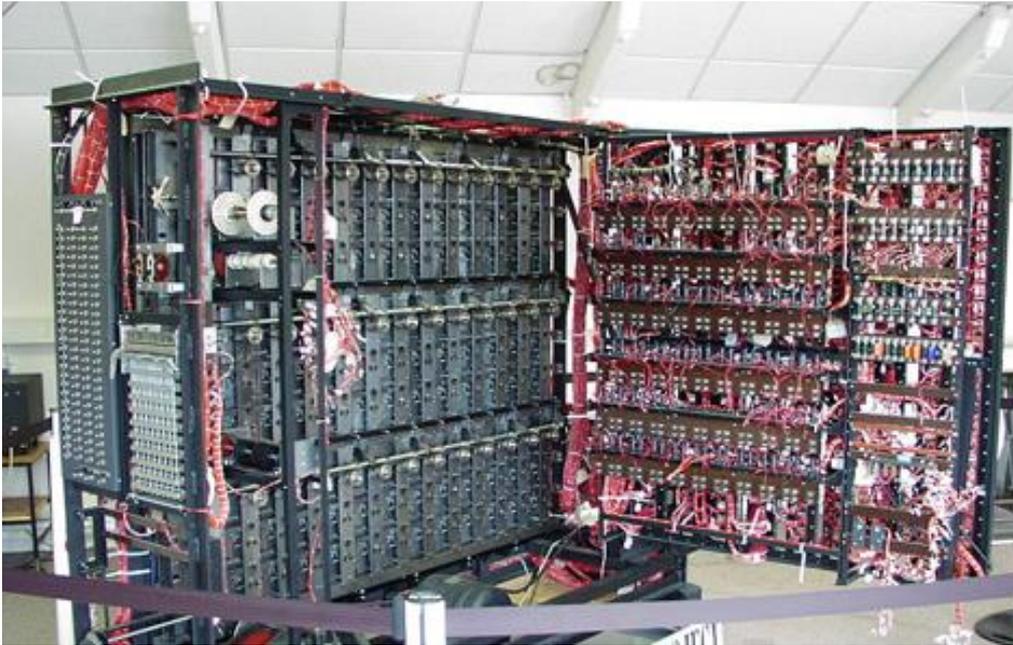
## 2.4 INTELIGÊNCIA ARTIFICIAL

De acordo com Pereira (2014), a Inteligência Artificial (IA) pode ser entendida como uma disciplina científica que utiliza capacidades de processamento de padrões da computação com o objetivo de encontrar métodos genéricos para automatizar atividades perceptivas, cognitivas e manipulativas através do computador. Além disso, a IA é sustentada por dois grandes pilares: a experimentação e a tecnologia, isso é, há investigação acompanhada de experimentação, mas também há uma vertente tecnológica que, quando combinadas com a investigação e experimentação, são capazes de promover uma revolução: a automatização de faculdades mentais por meio da modelização em computador.

Apesar de parecer recente o fato da Inteligência Artificial estar presente em vários dispositivos que utilizamos em nosso cotidiano, como programas de computador, aplicativos, ferramentas de diagnóstico médico, etc, a relação homem-máquina já se dá há mais de 60 anos – quando foi concebida com o objetivo de auxiliar durante a Segunda Guerra Mundial. Segundo Fernandes (2019), o desenvolvimento da tecnologia para fins militares mostrou-se uma peça fundamental durante os grandes conflitos da humanidade, como a primeira e Segunda Guerra Mundial, pois auxiliaram a cessar ou até mesmo anteciparem o fim dos conflitos. Um exemplo desse avanço foi a criação e desenvolvimento de armas incrementadas, utilização de aviões e submarinos e o uso de dispositivos de comunicação e interceptação. De acordo com Fernandes (2019), esses dispositivos foram associados às máquinas cifrantes, isto é, máquinas produtoras de códigos combinatórios, equipadas com rotores criptográficos, sendo um exemplo icônico disso a Máquina Enigma – que foi utilizada pelo governo alemão ainda nos anos 1920. No entanto, foi graças a utilização de tal equipamento que culminou na invenção do primeiro computador do mundo. Segundo Davies (2008), a operação Ultra, que nada mais era do que uma colaboração da inteligência militar britânica com engenheiros e matemáticos poloneses, conseguiu desenvolver um modelo mais poderoso do que o dos alemães, o que permitiu decifrar os códigos da Máquina Enigma. Em contrapartida, os alemães foram capazes de desenvolver um

segundo modelo, o *B-Schreiber*, tão sofisticado quanto o dos britânicos – mas que foi colocado por terra pelo britânico considerado o “pai da computação”, Alan Turing. De acordo com Yuge (2019), a *Bomb*, calculadora eletromecânica desenvolvida por Turing, incrementou a capacidade de decifração dos códigos da Máquina Enigma e levou à criação do *Colossus* – o primeiro dos computadores.

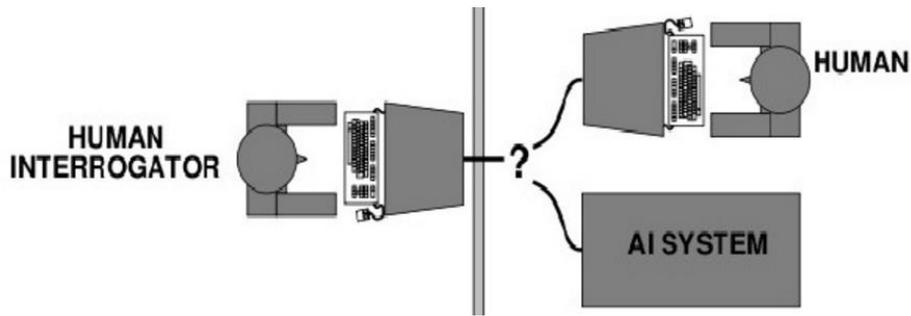
Figura 3 - Colossus, máquina desenvolvida por Alan Turing



Fonte: Fernandes (2019)

Alan Turing é considerado não só um dos pais da ciência da computação, mas como também da Inteligência Artificial. Conforme matéria publicada por Granatyr (2016), em 1950 o britânico criou um teste, Teste de Turing, que visava descobrir se uma IA é inteligente a ponto de enganar um humano, fazendo-o acreditar que se trata de um humano respondendo às suas perguntas através de textos. Para que um sistema de IA passe em um Teste de Turing, é necessário que 30% dos humanos consultados acreditem que quem está respondendo às perguntas se trata de um humano. A representação gráfica do Teste de Turing pode ser vista na figura a seguir:

Figura 4 - Representação gráfica do Teste de Turing



Fonte: Granatyr, J. 2016

A recente aceleração da inovação no campo da IA fez com que esse tema se tornasse recorrente em discussões governamentais, cotidiano de cidadãos comuns e até mesmo debates na mídia sobre as vantagens e consequências do avanço da Inteligência Artificial. De acordo com a Forbes (2017), existem quatro novas condições prévias que permitiram a aceleração da IA nos últimos anos, sendo elas:

- a) Tudo está conectado: Os sensores estão permitindo que coisas como edifícios, sistemas de transporte, máquinas, etc sejam conectadas através da nuvem, transformando-as em pequenos dispositivos que não só enviam dados, mas também recebem instruções (BOURCHARDT, E., 2017)
- b) Computação está se tornando barata: Novos chips estão chegando na mesma velocidade que seus predecessores e com a metade do custo. Isso significa que haverá um processador em tudo e que será possível colocar uma série de processadores baratos juntos em sistemas paralelos e distribuídos – a fim de obter a escala de computação necessária a um custo gerenciável para resolver problemas que antes eram impensáveis. (BOURCHARDT, E., 2017)
- c) Os dados estão se tornando o novo petróleo: Durante a revolução industrial, o petróleo era o que nutria as operações e, por isso, ter acesso a esse bem era uma vantagem competitiva imensurável. Hoje em dia, no entanto, são os dados que alimentam a revolução digital e, da mesma forma, as organizações que têm acesso exclusivo e que tem capacidade de processá-los da melhor forma é quem tem vantagem frente ao mercado (FORBES, 2017).
- d) *Machine Learning* está se tornando o novo motor de combustão: O aprendizado de máquina é uma maneira de usar algoritmos e modelos matemáticos para descobrir padrões implícitos em dados não refinados. As máquinas podem,

dessa forma, usar esses padrões complexos para descobrir por conta própria se um novo ponto de dados se encaixa ou é similar o suficiente para prever os resultados futuros. Já o *Deep Learning* permite aos computadores “ver” ou distinguir objetos e texto em imagens e vídeos (FORBES, 2017).

É irônico que a Inteligência Artificial tenha o potencial de fornecer informações em tempo real em qualquer lugar da organização, a qualquer momento, melhorar o desempenho e, mesmo assim, a maioria das empresas permaneçam amorfas diante de tal mudança. De acordo com o Data Science Academy (2018), para que os projetos de IA avancem e tenham sucesso dentro das organizações, é essencial que haja o apoio de executivos de nível C, ou seja, dos executivos seniores mais altos de uma companhia, como *Chief Executive Officer*, *Chief Financial Officer* e *Chief Operating Officer*, isto é, dos CEO, CFO e COO. Segundo Gartner (2019), isso significa que os executivos nível C devem assumir a responsabilidade e apoiar os projetos de desenvolvimento da Inteligência Artificial dentro de suas organizações – deixando claro que esse é um passo fundamental para o futuro dos negócios, maximizando a diversidade de times e fomentando a criatividade dentre departamentos.

Segundo o Data Science Academy (2018), implementar a IA com sucesso em uma empresa não é uma tarefa simples. Em uma pesquisa realizada pela Forbes (2019), 53% dos executivos acreditavam que seu mercado já havia enfrentado uma disrupção causada pela Inteligência Artificial. No entanto há ainda muitos equívocos que rodeiam essa nova Mega Tendência. Um deles é argumentar dizendo que a IA elimina funcionários: A empresa Amazon, por exemplo, revolucionou o mercado varejista ao oferecer aos seus clientes uma rede de supermercados inteligentes, a Amazon Go. De acordo com uma matéria publicada por Martin (2018), os clientes podem simplesmente entrar na loja, pegar o que quiserem da prateleira e sair da loja que, automaticamente, graças as ferramentas de IA implantadas (como o *Deep Learning* e sensores dentro da loja inteligente), o cliente é debitado diretamente do seu cartão de crédito cadastrado – evitando possíveis filas em caixas e otimizando o tempo do cliente. No entanto, as lojas ainda contam com o apoio de alguns funcionários para auxiliar os clientes a realizar funções como reposição de itens nas prateleiras, esclarecimento de qualquer dúvida, etc. Desta forma, pode-se dizer que a Inteligência Artificial não está eliminando funcionários, mas sim modelos de negócios (MCKENDRICK, J. 2018)

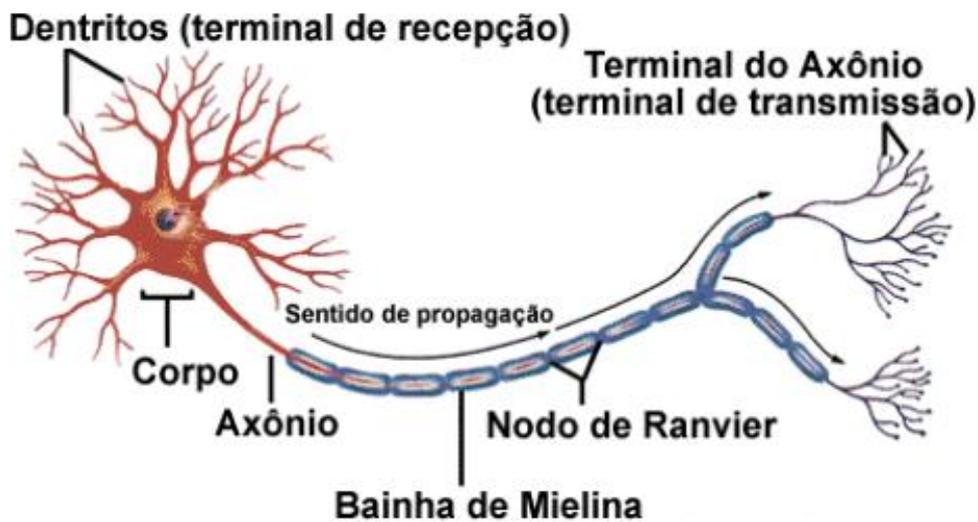
De acordo com Gartner (2019), o verdadeiro problema que acerca a IA dentro das grandes organizações é a criação de expectativas empresariais impossíveis de serem alcançadas por não existirem suficientes profissionais competentes para atendê-las, o que causa uma falta de criatividade e eficiência na concepção destes sistemas de IA. Porém, ferramentas como a adoção de linguagens de programação abertas, como o Python, e aumento da popularidade de IA possibilitaram que o conhecimento sobre as suas técnicas fossem mais difundidas dentro da organização e da sociedade - possibilitando a formação de um maior número de funcionários competentes ou engajados a aprender para atender as expectativas do mercado. A rápida difusão da informação, porém, de acordo com Ferreira (2018), implica em diversos dilemas éticos, como problemas de privacidade de dados, questões de segurança pública de máquinas autônomas executadas com IA, etc. Para pautar as preocupações éticas dentro das empresas e orientar tanto o pensamento quanto a rápida tomada de decisão, é essencial que as organizações desenvolvam uma estrutura interna de ética de IA. A Conferência de Asilomar, por exemplo, surgiu com o intuito de discutir e apresentar propostas para minimizar os riscos do progresso científico. O documento criado a partir dessa conferência pode ser consultado por qualquer cidadão, principalmente organizações, uma vez que elenca diversos princípios fundamentais para guiar líderes empresariais, formuladores de políticas governamentais e tecnólogos conforme o avanço da Inteligência Artificial. O documento, segundo Ferreira (2018), já foi assinado por mais de 3.000 pessoas relacionadas ao mundo tecno-científico, como por exemplo Stephen Hawking e Elon Musk.

#### **2.4.1 Redes Neurais**

O sistema nervoso é composto por diversas células nervosas extremamente complexas que são chamadas de neurônios e que se conectam de uma forma específica e precisa formando, assim, os circuitos neurais ou as redes neurais (KHAN ACADEMY, 2017). Estas redes complexas funcionam da mesma forma que um processador de informações onde os neurônios são unidades de processamento e são interligados uns aos outros por meio das sinapses neurais. De acordo com Haykin (1999), a complexidade destas redes é evidenciada pela quantidade de neurônios e sinapses que existem: estima-se que existam 100 bilhões de neurônios de tipos diferentes no cérebro, que realizam 100 trilhões de sinapses neurais.

Segundo Haykin (1999) e a Khan Academy (2018), o neurônio (Figura 5) diferencia-se das outras células por ser capaz de processar informações. A sua morfologia consiste no corpo do neurônio (soma), que é constituído por: um núcleo que dá todo o suporte metabólico para a célula; o axônio (fibra nervosa), que é um prolongamento grande e único que aparece no soma e realiza a condução do impulso nervoso para o próximo neurônio; pelos dendritos, que são prolongamentos menores em forma de ramificações que emergem do soma e do final do axônio e que são responsáveis pela comunicação entre os neurônios.

Figura 5 - Esquematização de um neurônio biológico



Fonte: Martins, L. (2019)

Quando um estímulo é disparado e a membrana do neurônio o recebe, ela se despolariza invertendo, assim, nas cargas elétricas: as cargas negativas ficam alocadas externamente e as positivas internamente. Segundo Vilela (2004), a partir do momento em que um ponto do nervo se despolariza, um impulso nervoso se propaga dele em todas as direções até atingir as extremidades da fibra.

De acordo com um dos artigos divulgados pela Khan Academy (2018), as sinapses químicas ocorrem quando um potencial de ação chega a um terminal pré-sináptico, o que causa a despolarização da membrana e faz com que um pequeno número de vesículas sinápticas secretem dentro da fenda sináptica um composto químico chamado de neurotransmissor, que atua nas proteínas receptoras do neurônio seguinte. Segundo Vilela (2004), isso acaba causando alterações na permeabilidade da membrana neural pós-sináptica, podendo, assim, excitá-lo, inibi-lo

ou modificar de alguma forma a sua intensidade. Pelo fato das sinapses poderem ser ajustadas através das atividades que elas participam, pode-se dizer que uma rede neural é capaz de “aprender”.

Ainda de acordo com Vilela (2004) e a Khan Academy (2018), os neurônios se comunicam através de uma sequência curta de pulsos e a mensagem é modulada na frequência destes pulsos. A frequência é um milhão de vezes mais lenta do que a maior velocidade de chaveamento de um circuito elétrico. Todavia, tarefas complexas são realizadas em poucos milissegundos devido a forma paralela do cérebro.

#### **2.4.1.2 Rede Neural Artificial**

Muito se tem discutido acerca de redes neurais artificiais (RNAs) - porém poucos sabem o que é. Segundo Braga, Ludermir e Carvalho (2000), RNAs são sistemas paralelos distribuídos e compostos por unidades de processamento simples que, por sua vez, tem a capacidade de calcular sistemas matemáticos não lineares e complexos. Esse sistema, que é baseado no cérebro humano, pode ser dividido em camadas e interligadas em diversas conexões - que estão relacionadas a pesos, que ponderam o modelo. A rede é extremamente atrativa para a resolução de problemas complexos devido a capacidade de aprender com uma quantidade baixa de exemplos e generalizar informações aprendidas, posteriormente dando respostas coerentes para dados não conhecidos.

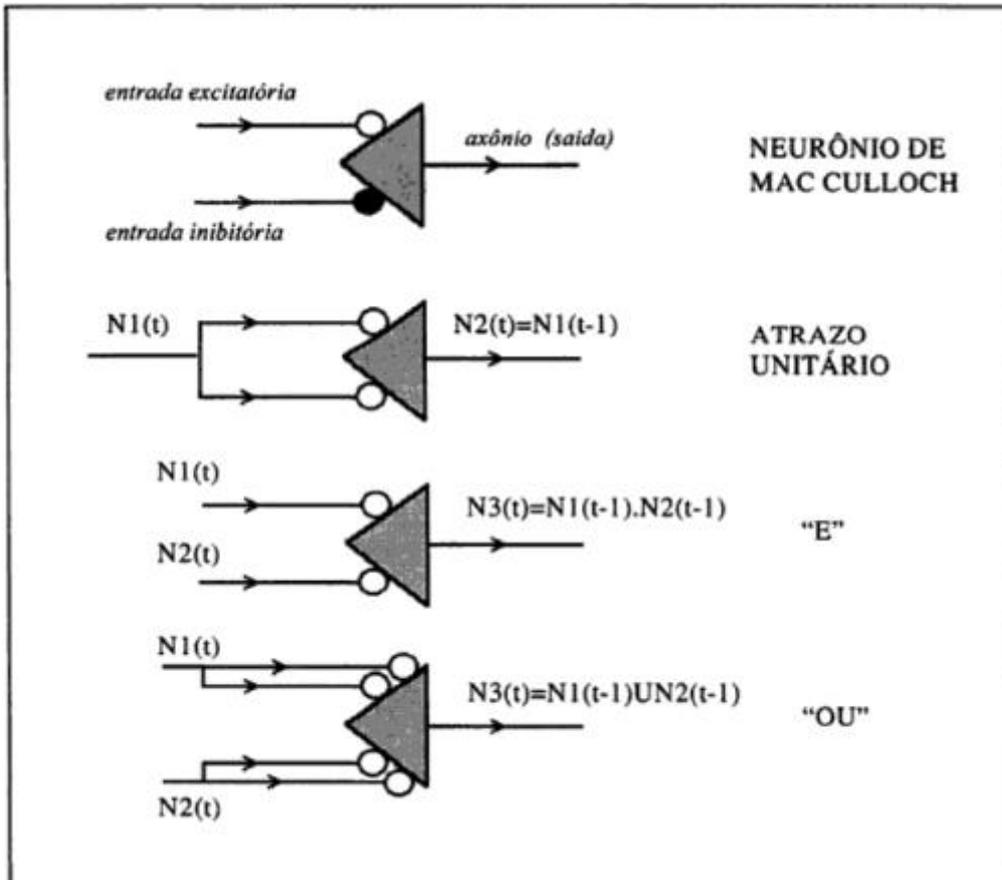
De acordo com o Data Science Academy (2019), sua concepção surgiu na década de 1940, por Warren McCulloch e Walter Pitts, através da publicação de um artigo. O trabalho consistia na semelhança de células nervosas vivas e um processo eletrônico binário. Recentemente, encontrasse inúmeras redes neurais artificiais muito mais complexas e diferentes, como por exemplo o Perceptron, Adaline e Madaline.

O processamento de uma rede neural é dividido em duas grandes etapas, sendo elas: aprendizagem e a real aplicação da rede. A aprendizagem é uma etapa em que as informações estão sendo aprendidas pela rede, para que se possa ajustar os devidos pesos e, assim, a rede aprender as novas informações que foram adicionadas; a aplicação da rede é onde se tem as respostas da rede em relação os dados apresentados a ela (CALDEIRA et al., 2007).

Segundo Kovács (2002), os pioneiros McCulloch e Pitts, com a rede neural relativamente simples, implementava máquinas booleanas no sistema nervoso em um

nível simbólico - que envolveu um sistema de uma grandeza maior do que o esperado pelos desbravadores. O sistema booleano proposto pode ser visto na Figura 6 abaixo:

Figura 6 - O Neurônio Booleano de McCulloth e implementações de algumas funções booleanas.



Fonte: Kovács (2002).

#### 2.4.1.2.1 Perceptron

Haykin (1999) explica que o Perceptron, modelo desenvolvido pelo americano Frank Rosenblatt e que pode ser visto na equação 24 a seguir, foi construído com base no modelo neural não-linear de McCulloth-Pitts. Rosenblatt propôs que um Perceptron – que pode ser entendido como um modelo matemático que recebe várias entradas e tem como resultado uma saída binária – deveria ter pesos associados, sendo estes números reais que expressam a importância das respectivas entradas para a saída. A saída, por sua vez, pode ser determinada pela soma ponderada, menor ou maior do que um valor limiar, o chamado *threshold*.

$$output = \begin{cases} 0 & \text{se } \sum_j W_j X_j \leq threshold \\ 1 & \text{se } \sum_j W_j X_j > threshold \end{cases} \quad (24)$$

Onde:

$X_j$ : é o valor de entrada  $j$

$W_j$ : é o peso da entrada associado a um valor de entrada  $j$

De acordo com o Data Science Academy (2019), a modelagem do Perceptron pode ser simplificada, transformando  $\sum_j W_j X_j$  em um produto – onde  $W$  e  $X$  são vetores cujos componentes são os pesos e entradas, respectivamente. Além disso, pode-se substituir o *threshold* pelo oposto do bias do Perceptron, isto é,  $b = -threshold$ . O bias, por sua vez, é a facilidade pela qual é possível obter uma saída 1 através do Perceptron, onde quanto mais negativo, mais difícil é de se obter esta saída. Desta forma, a fórmula pode ser reescrita como:

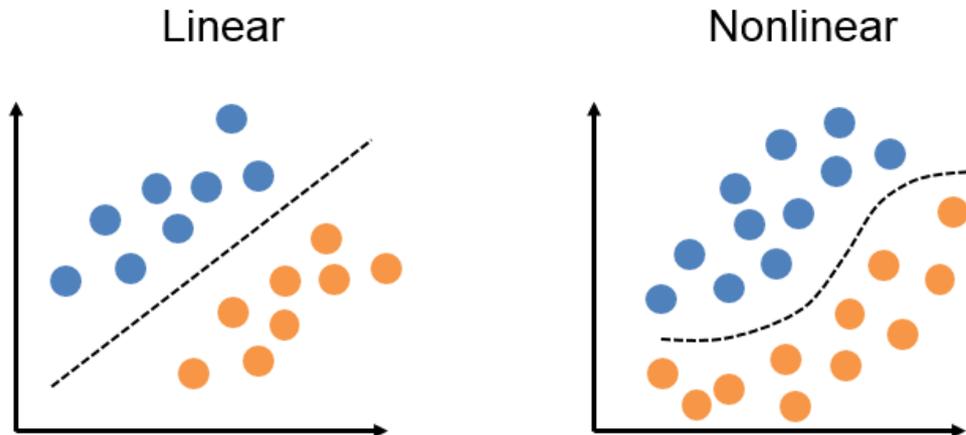
$$output = \begin{cases} 0 & \text{se } W \times X + b \leq 0 \\ 1 & \text{se } W \times X + b > 0 \end{cases} \quad (25)$$

Ainda segundo o Data Science Academy (2019), o Perceptron pode ser classificado de duas formas distintas: quando de camada única, é chamado de rede neural e quando o Perceptron possui várias camadas, é chamado de Rede Neural Artificial. O modelo pode ser treinado (*Machine Learning*) e esse processo consiste em ensinar os valores ideais de pesos e bias a partir do fornecimento de entradas e possíveis saídas. Com o modelo corretamente treinado, ao apresentar um novo conjunto de entradas, o modelo será capaz de prever a saída.

#### 2.4.1.2.1.2 Operações Lógicas e Regiões Linearmente Separáveis

Segundo o Data Science Academy (2019), ao considerar a função linear descrita na equação 25 acima, é possível traçar uma reta no plano cartesiano, separando-o em duas regiões distintas, isto é, regiões linearmente separáveis.

Figura 7 - Funções lineares ou não



Fonte: Data Science Academy (2019)

De acordo com o Data Science Academy (2019), um único Perceptron só é capaz de solucionar funções linearmente separáveis e, em casos que não seja possível gerar um hiperplano, é necessário utilizar mais de um Perceptron, isto é, uma Rede Neural Artificial. Essas funções, no entanto, são capazes de gerar condições lógicas, como AND, OR, NOT e XOR – sendo as três primeiras provenientes de funções linearmente separáveis e a última de funções não separáveis. Como é possível ver na Figura 8 a seguir, as entradas – que estão representadas nos eixos e que só podem assumir valores 0 ou 1, resultam em saídas diversas, conforme a condição lógica utilizada. A linha vermelha evidencia que lógicas como AND, OR ou NOT correspondem a funções linearmente separáveis.

Figura 8 - Funções Linearmente Separáveis

1	0	1	1	1	1	1	1	0
0	0	0	0	0	1	0	1	0
AND	0	1	OR	0	1	NOT	0	1

Fonte: Data Science Academy (2019)

Por outro lado, funções como XOR (Ou Exclusivo), fornece saídas de nível baixo (zero) quando todas as entradas têm o mesmo valor e, caso haja pelo menos uma entrada diferente entre si, a saída assume nível alto (um). Desta forma, não é possível separar o plano com apenas uma reta e, neste caso, deveria ser utilizada uma Rede Neural Artificial para solucionar o problema (Data Science Academy, 2019).

Figura 9 - Lógica XOR

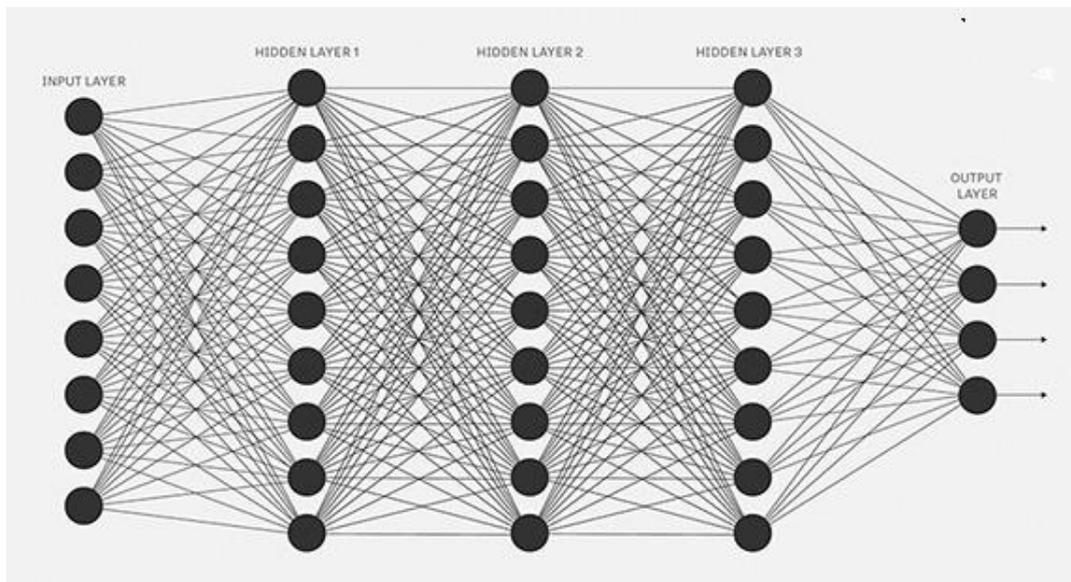
1	1	0
0	0	1
XOR	0	1

Fonte: Data Science Academy (2019)

#### 2.4.1.2.2 Função de Ativação

A Rede Neural é um mecanismo de aprendizagem de máquina e, assim como no funcionamento do cérebro, o estímulo é recebido e, a partir dele, um resultado é gerado. Conforme as atividades ficam mais complexas, mais neurônios são envolvidos no processo e, conseqüentemente, maior se torna a “rede de informações”. Abaixo é possível notar a semelhança de uma Rede Neural Artificial (mais de um Perceptron envolvido) com o processo que ocorre no cérebro humano:

Figura 10 - Modelo de Rede Neural Artificial



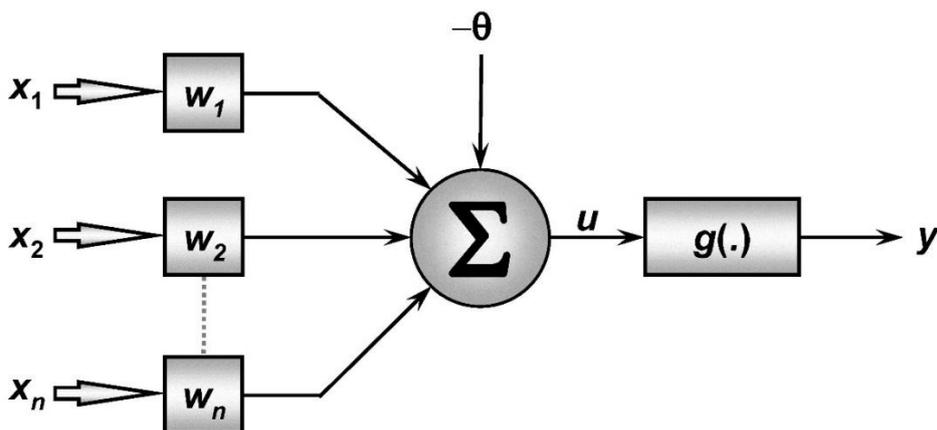
Fonte: Data Science Academy (2019)

Os neurônios, que são representados pelos pontos pretos da imagem acima, são caracterizados, de acordo com Romanelli (2017), por três elementos: o peso, bias e função de ativação. Como dito anteriormente, as camadas de entrada são alimentadas com os dados de entrada e estes dados são processados pelos neurônios – que realizam a transformação linear a partir dos pesos e bias. No entanto,

a transformação não linear, de acordo com o Data Science Academy (2019), é feita pela função de ativação. O fluxo de propagação da informação das camadas de entrada para camadas mais ocultas até o envio da saída final, é chamado de “movimento de propagação direta” e, caso o resultado não esteja de acordo com o esperado, os pesos e bias de cada neurônio são recalculados com base no erro, ou seja, uma *backpropagation*, para que só depois de validados os pesos e bias sejam usados para as projeções.

Segundo Sharma (2017), a função de ativação tem como objetivo principal ponderar as informações relevantes para o neurônio, isto é, filtrar quais informações devem ser consideradas e quais devem ser ignoradas – o que determina a ativação do neurônio. A função de ativação pode ser entendida como sendo mais uma camada matemática pertencente a Rede Neural, realizando uma transformação não linear e possibilitando que o modelo aprenda e execute tarefas mais elaboradas. A função de ativação é o que possibilita a *backpropagation*, como já mencionado, a partir do fornecimento dos gradientes e erros para atualizar os pesos e bias – fazendo com que essas pequenas mudanças não alterem drasticamente o output, como se fosse um “ajuste fino”. O modelo da rede com a função de ativação pode ser visto a seguir:

Figura 11 - Rede Neural com função de ativação



Fonte: Data Science Academy (2019)

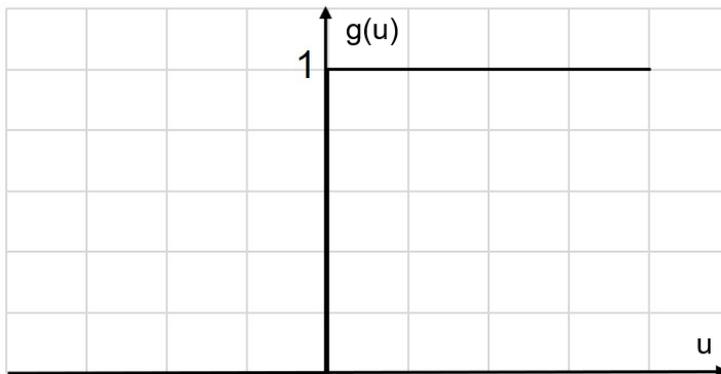
#### 2.4.1.2.2.1 Função de Etapa Binária

De acordo com Nielsen (2019), a função de etapa binária (ou Função Degrau) está relacionada ao valor limiar, o *threshold* e tem como saída apenas dois valores: um ou zero, isto é, saída alta ou baixa. Para que o valor de saída seja alto, o potencial

de ativação deve ser igual ou superior à zero e para que o valor de saída seja baixo, o potencial de ativação deve ser negativo – como é possível ver na equação 26 e Figura 12 a seguir:

$$g(u) = \begin{cases} 1 & \text{para } u \geq 0 \\ 0 & \text{para } u < 0 \end{cases} \quad (26)$$

Figura 12 - Gráfico Função Degrau



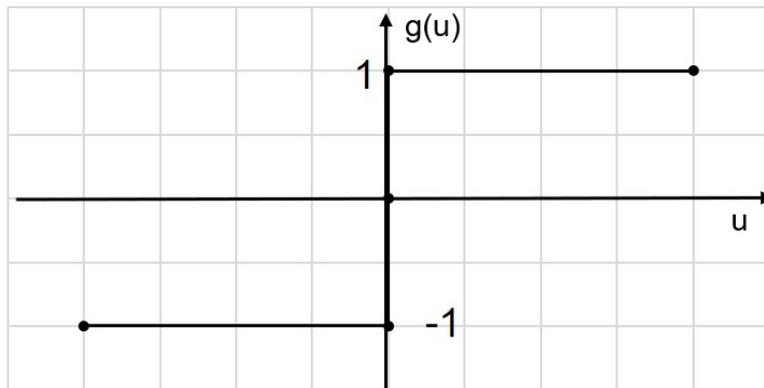
Fonte: Autores (2019)

#### 2.4.1.2.2 Função Degrau Bipolar

Ainda segundo Nielsen (2019), diferente da função degrau descrita na seção anterior, a Função Degrau Bipolar pode assumir três valores: 1, 0 ou -1. Para valores de ativação positivos, a saída assume valor alto (um); para valores nulos, a saída assume valor baixo (zero) e, em casos que a função de ativação assuma valores negativos, a saída será -1.

$$g(u) = \begin{cases} 1 & \text{para } u > 0 \\ 0 & \text{para } u = 0 \\ -1 & \text{para } u < 0 \end{cases} \quad (27)$$

Figura 13 - Gráfico Função Degrau Bipolar



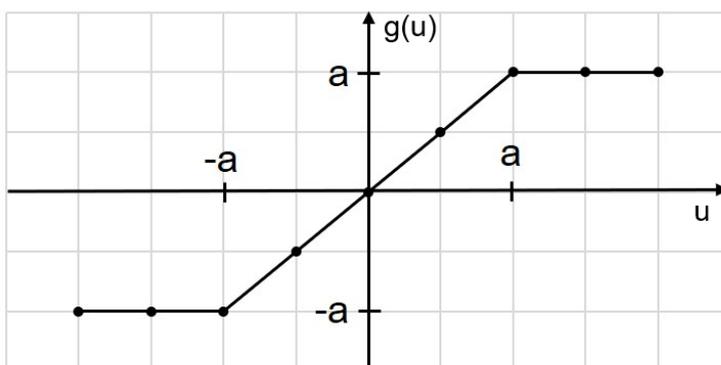
Fonte: Autores (2019)

#### 2.4.1.2.2.3 Função Rampa Simétrica

Nesse caso, de acordo com Finocchio (2014), os valores de saída são iguais aos valores dos potenciais de ativação – quando definidos entre o intervalo  $[-a, a]$  e, caso contrário, os valores se restringem aos limites. Dessa forma, temos que:

$$g(u) = \begin{cases} a & \text{para } u > a \\ u & \text{para } -a \leq u \leq a \\ -a & \text{para } u < -a \end{cases} \quad (28)$$

Figura 14 - Gráfico Função Rampa Simétrica



Fonte: Autores (2019)

#### 2.4.1.2.2.4 Função Sigmoide ou Logística

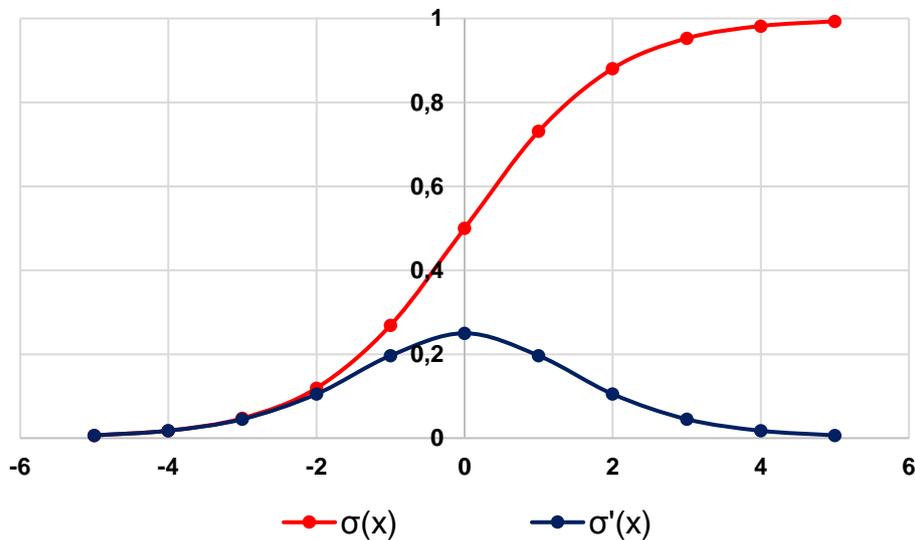
Essa função é recomendada para modelar comportamentos como os dos neurônios biológicos, uma vez que estes funcionam de forma binária (ativo ou não ativo). De acordo com Costa (2017), a função sigmoide é uma função real, diferenciável e limitada que é definida para todos os valores reais de entrada e seu valor de saída pertence ao intervalo  $(0, 1)$  - possuindo derivada positiva em cada ponto:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (29)$$

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (30)$$

A representação gráfica da função sigmoide pode ser vista na Figura 15 a seguir:

Figura 15 - Função de ativação Sigmoide



Fonte: Autores (2019)

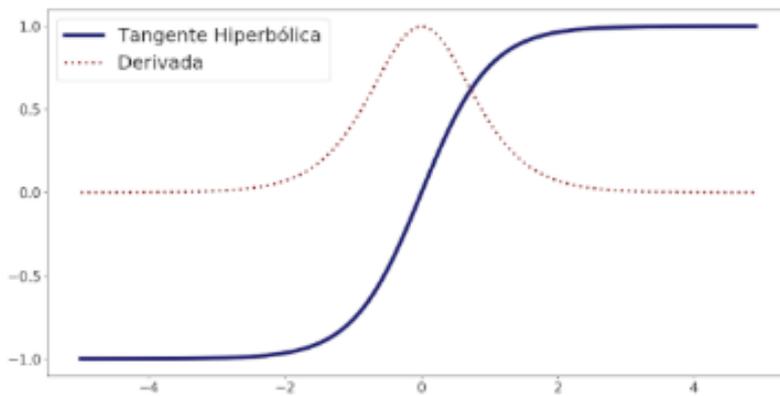
A representação gráfica acima nos permite notar que a derivada da função sigmoide satura e, por esse motivo, de acordo com Costa (2019), derivadas menores que 1 e tendendo a zero fazem com que a propagação do gradiente se dissipe nessas regiões – o que causa dificuldade no treinamento.

#### 2.4.1.2.2.5 Função Tangente Hiperbólica

Para Nielsen (2019), assim como a função sigmoide, a função de ativação tangente hiperbólica também é uma função real, diferenciável e limitada que é definida para todos os valores reais de entrada. Porém é comumente usada em redes neurais cujas saídas devem ser entre -1 e 1. A equação é dada abaixo, assim como o gráfico da função:

$$\phi(x) = \tanh(x) \quad (31)$$

Figura 16 - Função de ativação tangente hiperbólica



Fonte: Costa (2019)

A derivada dessa função é positiva, no entanto, é possível observar que apesar de ainda existirem algumas regiões de saturações, como na função sigmoide, o valor máximo da derivada é igual a 1. Por conta disso, Costa (2019) afirma que quando uma função sigmoide precisa ser utilizada, é recomendado a função tangente hiperbólica e não a função sigmoide.

#### 2.4.1.2.2.6 Função de Ativação Linear

Segundo Reis (2016), a função de ativação linear é a função mais básica pois não altera a saída de um neurônio e geralmente é utilizada nas camadas de saída em redes neurais de regressão. A equação e a representação gráfica são dadas abaixo:

$$\phi(x) = x \quad (32)$$

Figura 17 - Função de ativação Linear



Fonte: Autores (2019)

#### 2.4.1.2.3 Arquitetura de Redes Neurais Artificiais

De acordo com Finocchio (2014) e o Data Science Academy (2019), a arquitetura de redes neurais pode ser classificada em três categorias distintas conforme sua estrutura, sendo elas: estática, dinâmica ou *fuzzy* e pode conter camada única ou múltipla. Além disso, as arquiteturas podem ser diferenciáveis conforme o fluxo de informações ou conforme a arquitetura em que foram implementadas. Ainda segundo Finocchio (2014), o fluxo entre os neurônios pode ser estritamente no sentido de ida, no sentido de ida e volta, lateralmente conectadas, topologicamente ordenadas ou híbridas – enquanto que a arquitetura em que foram implementadas pode ser conforme características de seus nós, regras de treinamento e tipos de modelos de neurônio empregado.

##### 2.4.1.2.3.1 Redes Neurais Feed-Forward

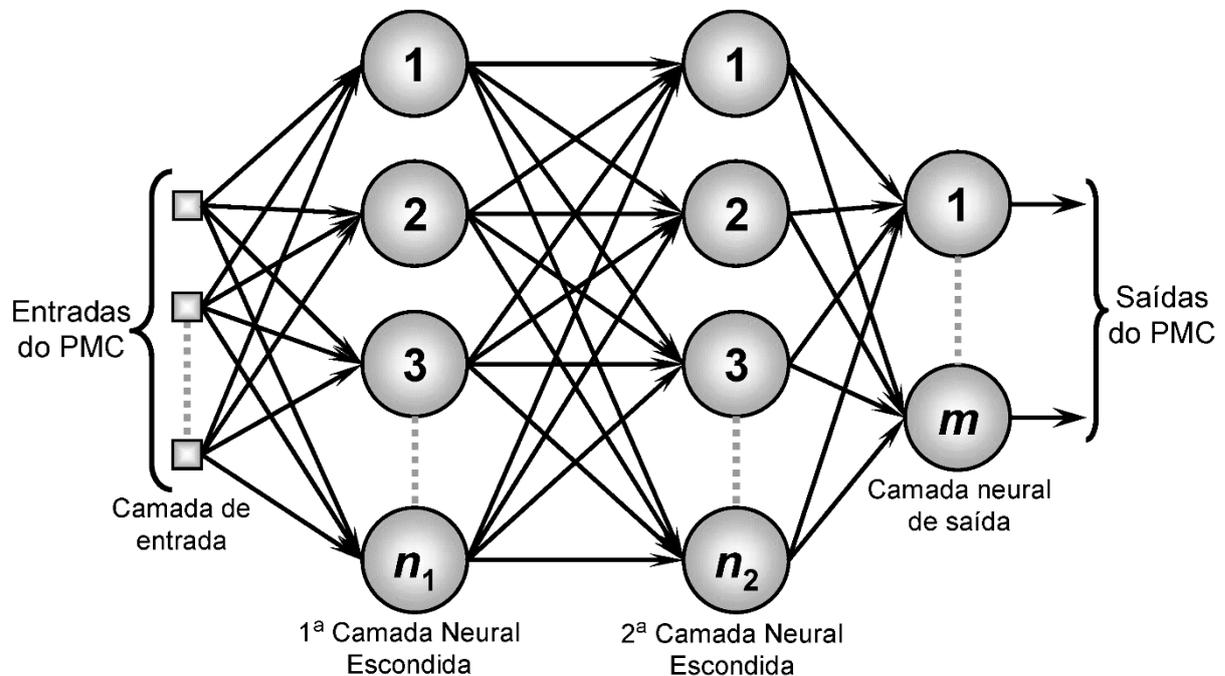
Sendo considerada a estrutura mais comum em aplicações práticas, a rede feed-forward consiste em uma primeira camada de entrada e a última de saída e há casos em que há camadas intermediárias, isso é, camadas “ocultas” ou “profundas” (deep learning). De acordo com Nielsen (2019) e Finocchio (2014), esse tipo de estrutura é capaz de calcular uma série de transformações que alteram as

semelhanças entre os casos e as atividades dos neurônios em cada camada são uma função não-linear das atividades na camada anterior.

#### 2.4.1.2.3.2 Redes Neurais Recorrentes

Apesar de ser biologicamente mais realistas, as estruturas recorrentes não são simples de serem treinadas. Isso se deve ao fato de que o fluxo de informação pode ocorrer em “loop”, ou seja, o fluxo de informação não é unidirecional: algumas saídas são utilizadas como novas entradas dos neurônios, emitindo saídas atuais com base em saídas anteriores. De acordo com Finocchio (2014), esses modelos “são equivalentes a redes muito profundas com uma camada oculta por fatia de tempo; exceto que eles usam os mesmos pesos em cada fatia de tempo e recebem entrada em cada fatia”. Dentre as principais redes que utilizam essa estrutura, podemos citar a Rede Hopfield e a Rede Perceptron Multicamadas (PMC) – a qual pode ser vista na Figura 18 abaixo e, em termos de aplicação, pode-se destacar as séries temporais, controle de processos e otimização.

Figura 18 - Rede Perceptron de Multicamadas



Fonte: Moreira (2018)

#### 2.4.1.2.3.3 Redes Neurais Conectadas Simetricamente

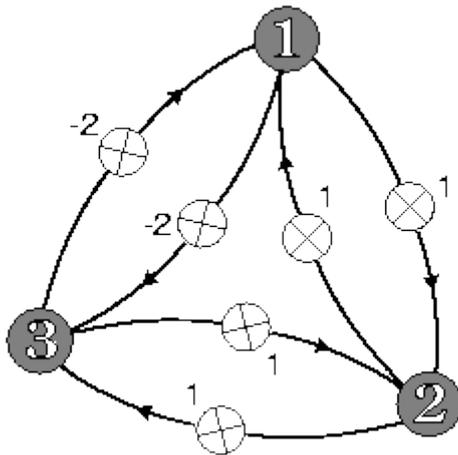
Esse tipo de estrutura se assemelha a estrutura retroalimentada citada na seção anterior, diferenciando-se apenas pelo fato de que as conexões entre as unidades possuem o mesmo peso em ambas as direções (simétricas). Segundo Wittek (2017) e Finocchio (2014), as redes neurais conectadas simetricamente são mais simples de serem analisadas do que as redes retroalimentadas. As redes com essa estrutura e que não possuem unidades ocultas, são chamadas de Redes Hopfield enquanto as que possuem unidades ocultas são chamadas de Máquinas de Boltzmann.

##### 2.4.1.2.3.3.1 Rede Hopfield

De acordo com Sousa (2014), um dos primeiros modelos de Rede Neural Artificial propostos foi o *Perceptron*, como mencionado na seção 5 deste trabalho, e era utilizado para classificar apenas padrões linearmente separáveis. Este modelo evoluiu para o *Perceptron* Multicamadas, que permitiu o aprendizado *error back propagation* – possibilitando que o modelo classifique qualquer tipo de padrão e não apenas os linearmente separáveis. No entanto, o aprendizado deste tipo de modelo era extremamente demorado e complexo, o que levou ao surgimento de novos modelos, permitindo aprimorar a eficiência computacional e até mesmo se assemelhar ao comportamento cerebral humano – como mencionado na seção anterior. Segundo Sousa (2014) e Wittek (2017), em 1982 o físico J. J. Hopfield apresentou um modelo recorrente que fazia analogia a um sistema físico real. A rede de Hopfield é organizada em uma única camada, na qual cada neurônio realiza sinapses com todos os demais, exceto ele mesmo, de forma recorrente (a saída de um neurônio será o valor de entrada dos demais). Em suma, de acordo com Sousa (2014), isso significa que uma dada resposta irá depender sempre do estado do sistema no intervalo de tempo anterior, conferindo à rede uma espécie de memória. Segundo Wittek (2017), um padrão de memória é armazenado de forma distribuída e pode ser recuperado com base no conteúdo total ou parcial do mesmo, isto é, o modelo funciona como uma memória associativa, semelhante ao que ocorre no cérebro humano. A esquematização da Rede de Hopfield pode ser visto da Figura 19, onde os círculos mais escuros representam os nós (neurônios) e os círculos brancos com cruzes

representam os pesos dos arcos. É importante notar que além de ser um modelo recorrente, os pesos dos arcos deste modelo em específico, que ligam os mesmos neurônios, possuem o mesmo peso – o que confere a característica de Rede Neural conectada simetricamente.

Figura 19 - Rede recorrente de Hopfield

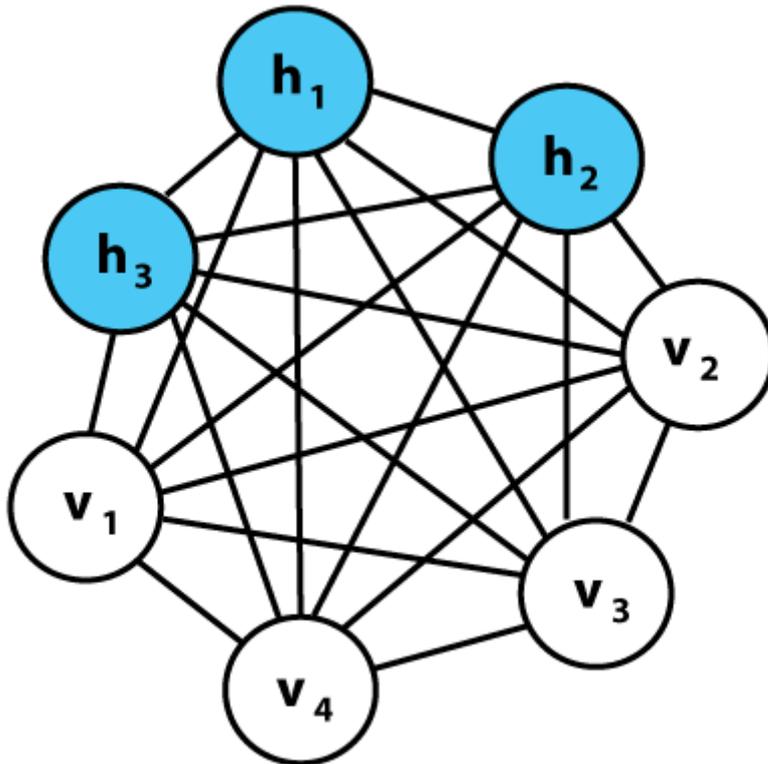


Fonte: Oliveira, et. al. (2003)

#### 2.4.1.2.3.3.2 Máquinas de Boltzmann

Diferente dos modelos citados anteriormente, as Máquinas de Boltzmann se diferem pelo fato de que possuem múltiplas direções de transmissão das informações (NIELSEN, M. 2019), isso é, todos os neurônios são conectados entre si e as informações fluem em todas as direções. Por este motivo, podemos diferenciar as Máquinas de Boltzmann, também, pelo fato do modelo não possuir camada de saída, apenas camada visual e camadas ocultas. O termo “camada de entrada” não foi utilizado propositalmente, uma vez que as Máquinas de Boltzmann não possuem dados de entradas propriamente; o modelo é considerado um Modelo Generativo Profundo, ou seja, a rede pode ser entendida como um sistema capaz de gerir suas próprias informações (DATA SCIENCE ACADEMY, 2019). Esse modelo, assim como o de Hopfield, é uma rede neural recorrente em que os nós tomam decisões binárias com algum viés. Abaixo é possível ver a esquematização do modelo de Boltzmann, onde as camadas brancas são as camadas visuais (*v*, *visual*) e as azuis são as camadas ocultas (*h*, *hidden*).

Figura 20 - Máquina de Boltzmann



Fonte: Sharma (2018)

Para ilustrar este modelo bastante complexo, Thano (2017) relaciona o modelo de Boltzmann com uma festa, onde a camada visual representa as pessoas que conhecemos e a camada oculta as pessoas que não conhecemos. Cada nó dessas camadas pode assumir valor alto ou baixo, isso é, 1 ou 0 – onde 1 representa o fato da pessoa ir à festa e 0 representa o contrário. A probabilidade de uma pessoa ir à festa depende da probabilidade de todas as outras irem também, isso é, caso uma pessoa X expresse a intenção de ir a festa e a pessoa X seja amiga da pessoa Y, a probabilidade da pessoa Y também ir à festa é alta. Agora, caso a pessoa X mantenha a decisão de ir, mas tem uma inimidade com a pessoa Z, as chances da pessoa Z não atender o evento são altas. Dessa forma, a probabilidade de uma pessoa ir à festa depende dessas conexões. De acordo com Thano (2017), pode ser ainda que algumas conexões assumam valor zero, indicando indiferença entre os elementos – o que no caso da festa seria o mesmo que dizer que a pessoa X e W são indiferentes entre si, não exercendo influência na decisão umas das outras. Ainda segundo o mesmo autor, há um estado da festa em que a maioria das pessoas presentes não se

gostam. Esse estado é de bastante tensão e há uma tendência para que ele não ocorra frequentemente. As Máquinas de Boltzmann, por sua vez, capturam este cenário e atribuem baixa probabilidade em estados com muita energia.

#### *2.4.1.2.4 Treinamento de Redes Neurais*

As Redes Neurais, como já descrito neste trabalho, são compostas por elementos não lineares, que operam em paralelo, e são classificadas de acordo com padrões ligados à biologia (TATIBANA, KAETSU, 2019). Sua característica mais importante é a capacidade e habilidade de aprender por intermédio de seu ambiente e, com isso, melhorar seu desempenho. Isso só é possível por conta do processo iterativo de ajustes aplicado a seus pesos – recalculando os valores sempre que necessário, como mencionado nas seções anteriores, e finalizando o processo de treinamento quando a rede neural consegue generalizar soluções para uma classe de problemas.

Os conhecimentos das redes neurais são armazenados nas sinapses, ou seja, nos pesos que são atribuídos a cada uma das conexões entre os neurônios. De acordo com Tatibana e Kaetsu (2019), estima-se que de 60 a 90% do total dos dados devem ser separados para o treinamento da rede, sendo que estes dados devem ser separados de forma aleatória para que a rede aprenda as regras associadas ao processo. Os outros 40 a 10% dos dados só são apresentados à rede na fase de testes para verificar se as regras geradas são capazes de produzir saídas condizentes.

##### *2.4.1.2.4.1 Aprendizado de Redes Neurais*

Um algoritmo de aprendizado é desenvolvido com base em um conjunto de regras com o intuito de solucionar um problema. Há uma vasta variedade de algoritmos de aprendizado sendo que a principal diferença entre eles é a maneira pela qual os pesos das sinapses são modificados. Esse processo de aprendizado pode ser dividido em duas categorias: Supervisionado ou Não Supervisionado.

#### 2.4.1.2.4.1.1 *Aprendizado Supervisionado*

Esse processo de aprendizagem consiste em gerar um conjunto de saídas a partir de um conjunto de entradas fornecido. De acordo com Finocchio (2014), quando um vetor de entrada é aplicado, a saída da rede é calculada e comparada com o respectivo padrão de saída. O erro, isto é, a diferença entre a saída gerada e a saída esperada, é então calculado e propagado no sentido contrário, ou seja, uma *backpropagation*. Dessa forma, segundo Nielsen (2019), os pesos são alterados em conformidade com algum algoritmo que tende a minimizar esse erro. Esse processo ocorre inúmeras vezes para cada vetor, como um método iterativo, e acaba quando o erro atinge um nível aceitável para o conjunto de treinamento.

#### 2.4.1.2.4.1.2 *Aprendizado Não Supervisionado*

Diferente do método anterior, esse método de aprendizado não requer que um agente externo indique a resposta desejada para que haja o recálculo dos pesos. Nesse caso, a rede se auto organiza, se utilizando de exemplos de situações análogas. Segundo Finocchio (2014), o algoritmo de treinamento “modifica os pesos da rede para produzir vetores de saída que são consistentes, isto é, vetores do conjunto de treinamento que são similares entre si produzirão o mesmo padrão de saída”. Espera-se, de acordo com o Data Science Academy (2019), que esse tipo de modelo de aprendizagem descubra as características particulares da população de entrada para que, então, produza saídas coerentes ao que se deseja. Dessa forma, é evidente a diferença entre o aprendizado supervisionado e o não supervisionado: enquanto o primeiro classifica os padrões em categorias pré-determinadas, o segundo desenvolve sua própria representação do estímulo de entrada.

### 3. TRABALHO DE CAMPO

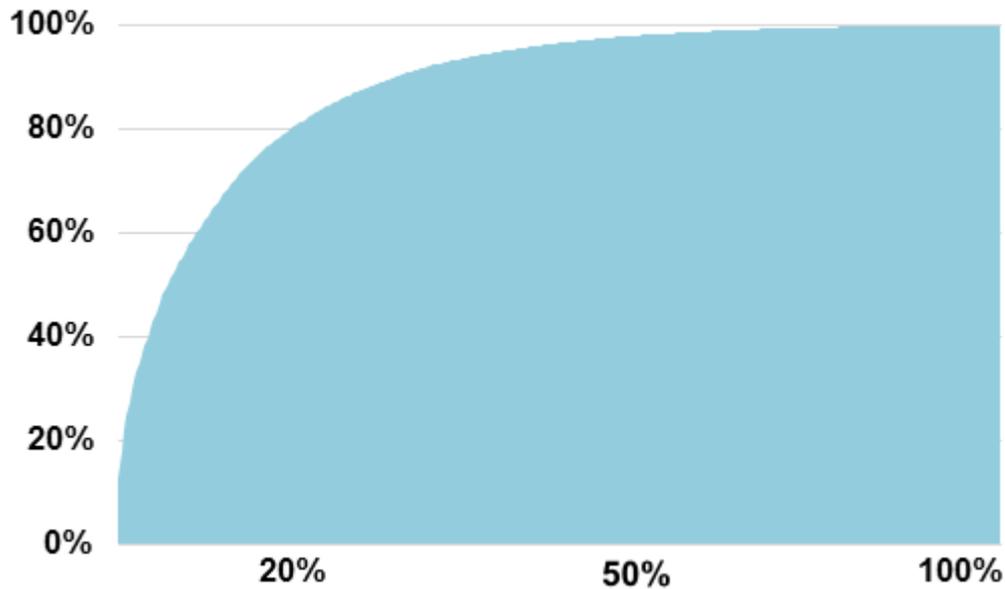
Os modelos de projeção de demanda citados ao longo do presente trabalho foram aplicados a uma empresa de pequeno porte de componentes eletrônicos, localizada em Mogi Mirim (São Paulo), e foram realizadas a partir do histórico de vendas de 5 anos do “Produto L”. O produto em questão foi escolhido com base na análise de Curva ABC – considerando a participação de todos os produtos do portfólio da empresa. Através do software estatístico Stata, foram elaborados resultados gráficos de comparação, assim como o cálculo de erros percentuais para verificar qual o método mais preciso para o caso em estudo. Já para as projeções gráficas de Redes Neurais, foi utilizado o Software SPSS, da IBM.

Para que o Stata gere as representações gráficas de acordo com os métodos estatísticos que se deseja estudar, é necessário inserir os algoritmos matemáticos para que o software seja capaz de calcular, por exemplo, a projeção através da média móvel e seu respectivo erro percentual associado a essa previsão. Além disso, a partir do input da base de dados, que contém o histórico de vendas de 60 meses (5 anos), foi utilizado 80% do banco de dados (48 meses) para realizar a projeção e 20% para teste (12 meses) - como forma de comparar os resultados gerados com os realizados. A partir dessa segmentação do banco de dados, foi calculado o erro percentual da projeção através do MAPE (*Mean Absolute Percentage Error*) e plotado o gráfico com os resultados. Na seção Anexo no final do presente documento, é possível verificar as janelas geradas pelo Software Stata com as saídas projetadas a partir do input do histórico de vendas e do método estatístico da média móvel. De forma análoga foram utilizados os outros métodos estatísticos e calculados os respectivos erros percentuais.

#### 3.1 CURVA ABC

A partir do portfólio da empresa de componentes eletrônicos de Mogi Mirim, foi realizada uma análise da representação dos produtos no faturamento total e, através deste método, foi possível selecionar o produto para o qual os modelos de projeção de demanda seriam aplicados, como é possível verificar na representação gráfica e tabela a seguir:

Figura 21 - Curva ABC aplicada a empresa escolhida



Fonte: Autores (2020)

Tabela 1 - Ranking dos Top 5 produtos que tem maior participação no faturamento da empresa

<b>Produto</b>	<b>Participação no faturamento da Empresa</b>
Produto L	12,86%
Produto K	5,04%
Produto J	4,39%
Produto H	2,69%
Produto G	2,59%

Fonte: Autores (2020)

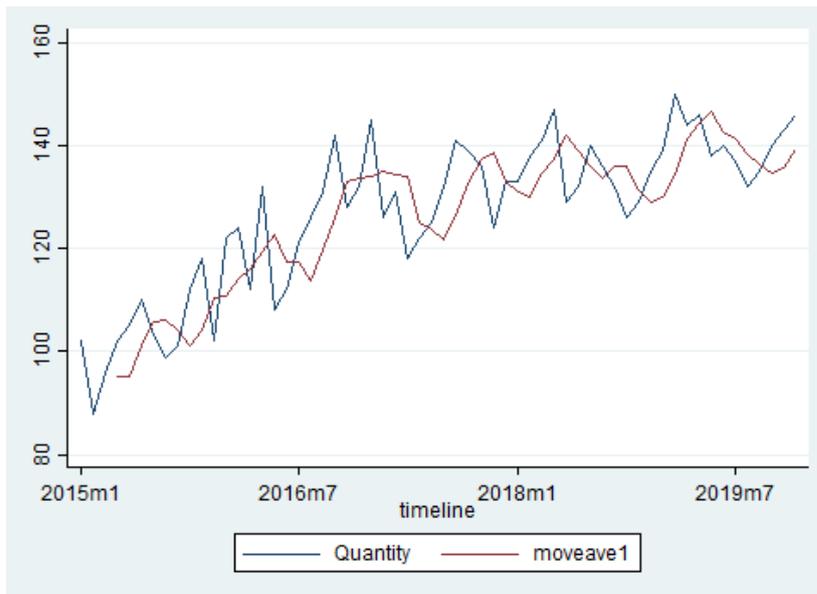
### 3.2 SAÍDAS DOS DIFERENTES MODELOS DE PROJEÇÃO DE DEMANDA

Uma vez inseridos os códigos dos modelos matemáticos que se deseja utilizar, e tendo uma base de dados adequada para sua utilização no Software Stata, correm-se um por um os métodos para obter seu output - com os gráficos e seus respectivos erros

### 3.2.1 Saída da Projeção Utilizado Média Móvel Simples

Para este modelo matemático, o erro MAPE é de 5,77% para a projeção dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método “Média Móvel Simples”.

Figura 22 - Saída Gráfica do Método Média Móvel Simples

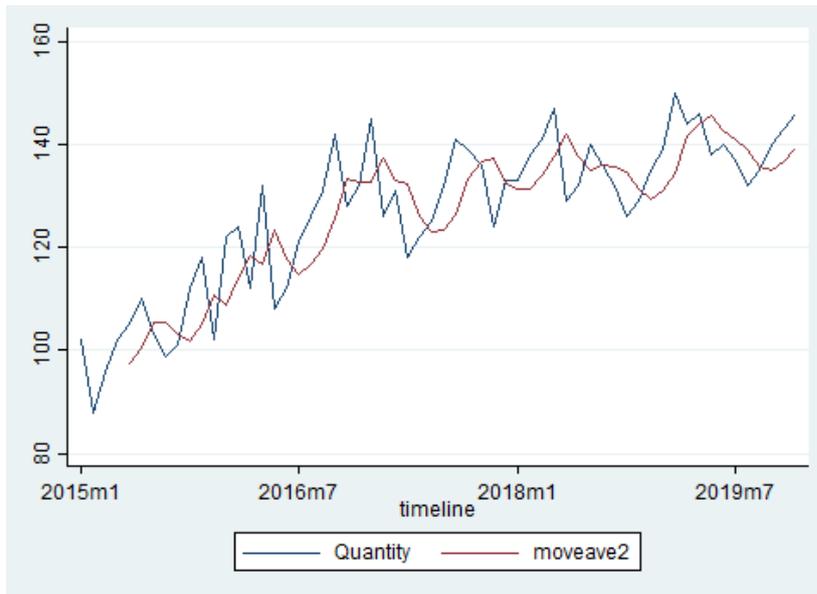


Fonte: Autores (2020)

### 3.2.2 Saída da Projeção Utilizando Média Móvel Ponderada

Para este modelo matemático, o erro MAPE é de 5,64%, para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método “Média Móvel Ponderada”.

Figura 23 - Saída Gráfica do Método Média Móvel Ponderada

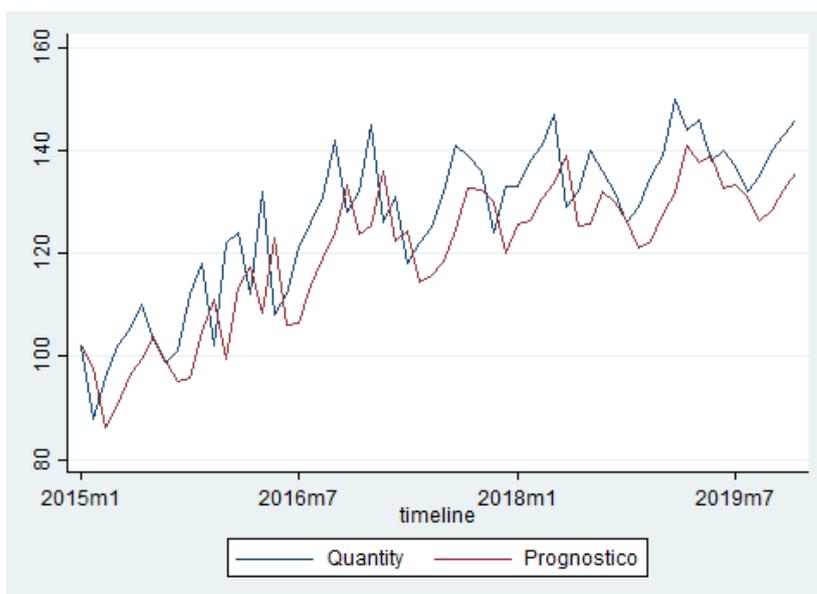


Fonte: Autores (2020)

### 3.2.3 Saída da Projeção Utilizando Amortecimento Exponencial Simples

Para esse modelo matemático, o erro MAPE é de 7,43% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método “Amortecimento Exponencial Simples”.

Figura 24 - Saída Gráfica do Método Amortecimento Exponencial Simples

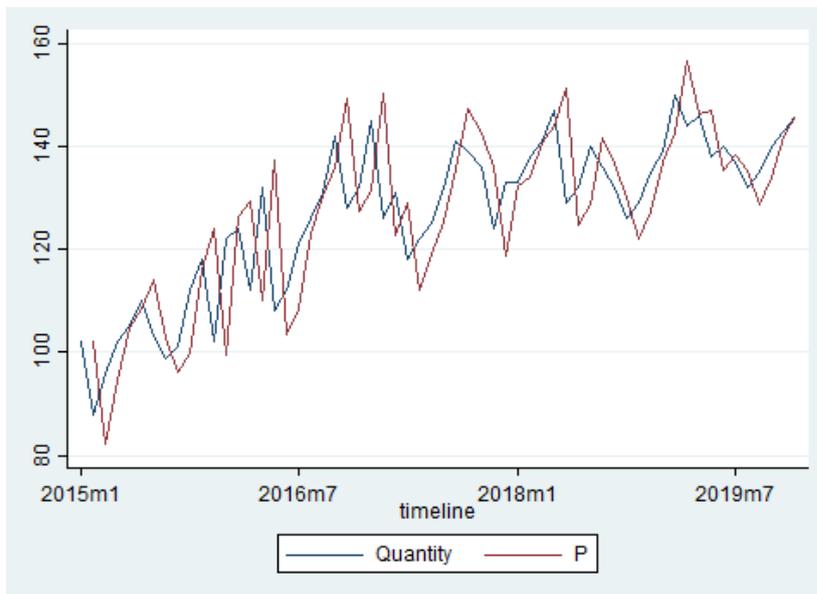


Fonte: Autores (2020)

### 3.2.4 Saída da Projeção Utilizando Amortecimento Exponencial Dupla

Para esse modelo matemático, o erro MAPE é de 6,97% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método “Amortecimento Exponencial Dupla”, ou também conhecido como “Método de Brown”.

Figura 25 - Saída Gráfica do Método Amortecimento Exponencial Dupla

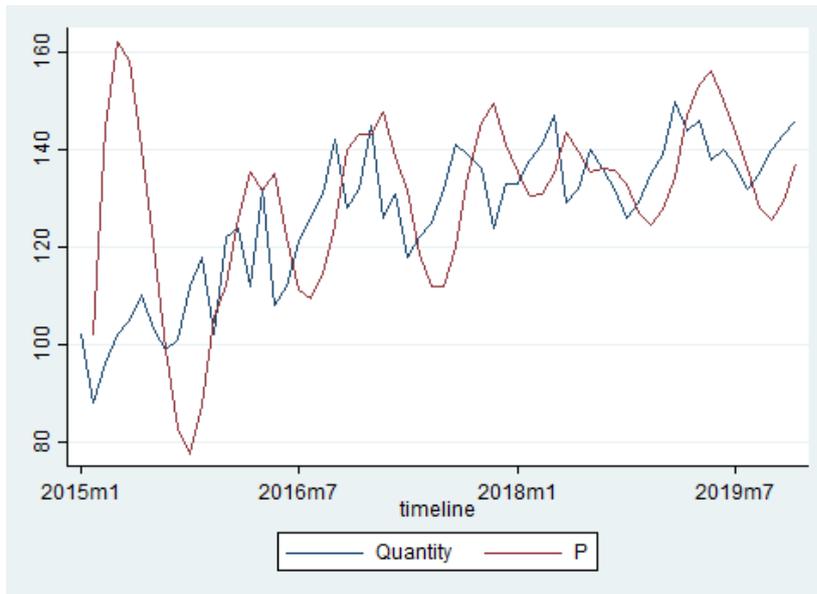


Fonte: Autores (2020)

### 3.2.5 Saída da Projeção Utilizando Suavização Exponencial Com Tendência

Para essa projeção, o erro MAPE é de 11,68% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método “Suavização Exponencial com Tendência”, ou também conhecido como “Método de Holt”.

Figura 26 - Saída Gráfica do Método Suavização Exponencial Com Tendência

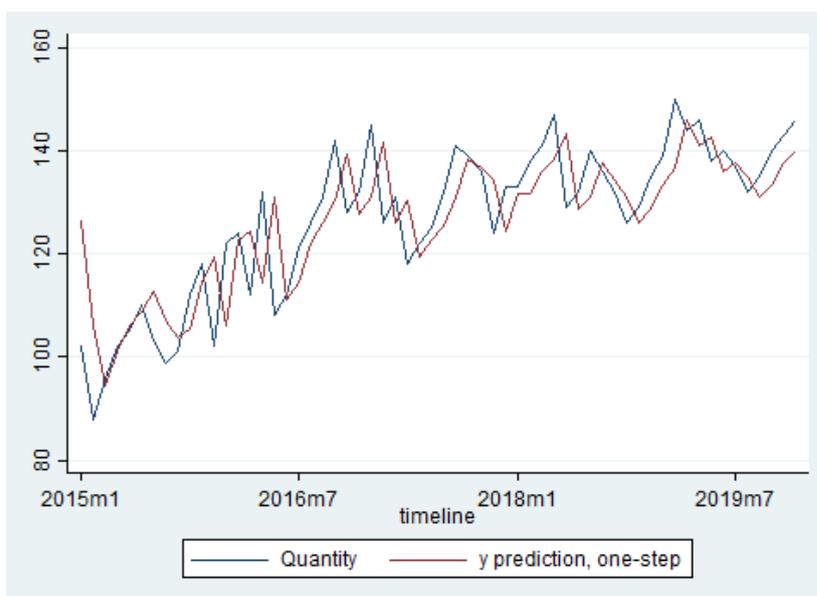


Fonte: Autores (2020)

### 3.2.6 Saída da Projeção Utilizando Método Ar(1)

Para esse método matemático, o erro MAPE é de 5,79% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método Arima 1.

Figura 27 - Saída Gráfica do Método Arima (1)

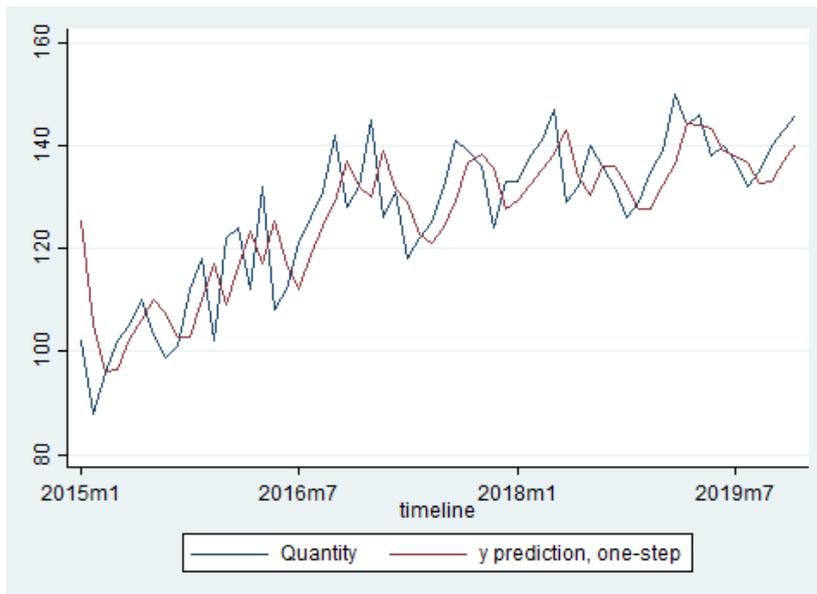


Fonte: Autores (2020)

### 3.2.7 Saída da Projeção Utilizando Método Arima (2)

Para esse modelo matemático, o erro MAPE é de 5,80% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método Arima 2.

Figura 28 - Saída Gráfica do Método Arima(2)

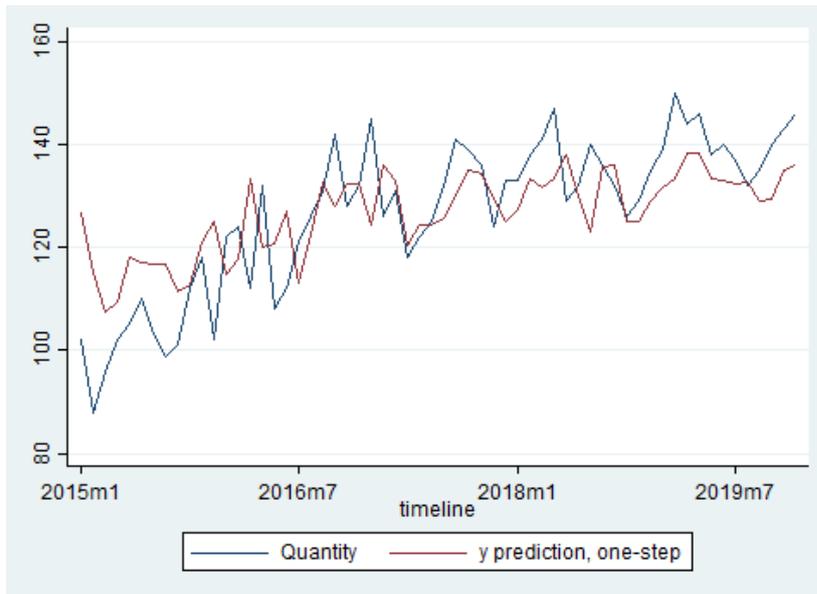


Fonte: Autores (2020)

### 3.2.8 Saída da Projeção Método Média Móvel 1 - Ma(1)

Para esse modelo matemático, o erro MAPE é de 8,04% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método Arima 1 Média Móvel.

Figura 29 - Saída Gráfica do Método Ma(1)

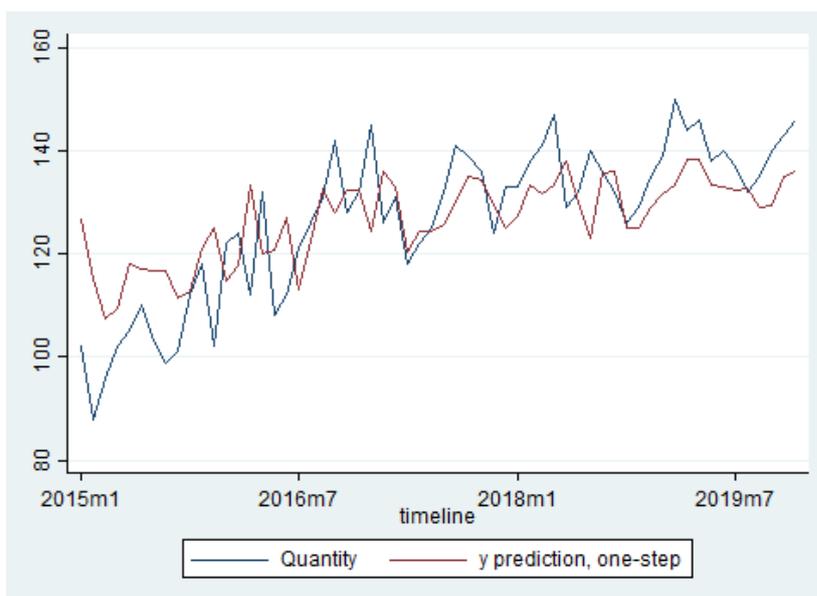


Fonte: Autores (2020)

### 3.2.9 Saída da Projeção Método Média Móvel 2 - Ma(2)

Para esse modelo matemático, o erro MAPE é de 7,04% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método Arima 2 Média Móvel.

Figura 30 - Saída Gráfica do Método Ma(2)

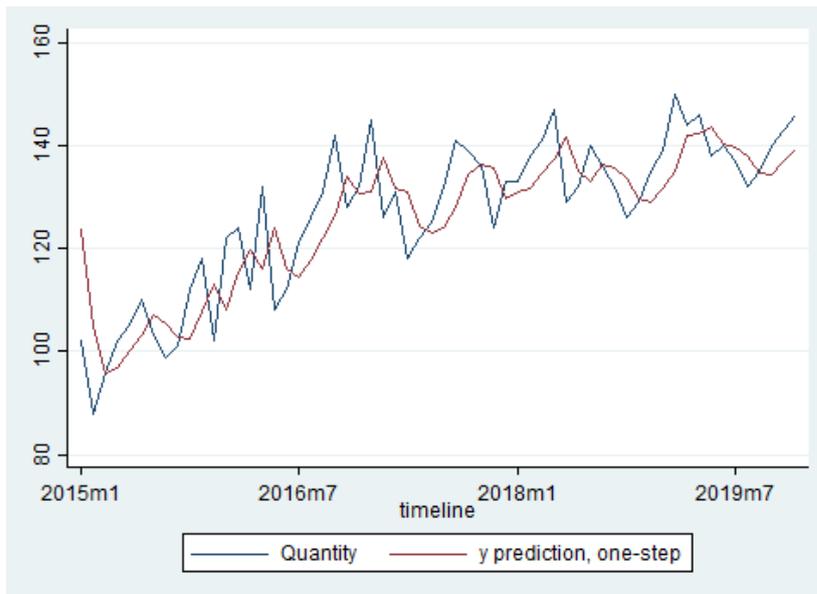


Fonte: Autores (2020)

### 3.2.10 Saída da Projeção Método ARMA(1,1)

Para esse modelo matemático, o erro MAPE é de 5,72% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método ARMA (1,1).

Figura 31 - Saída Gráfica do Método ARMA(1,1)

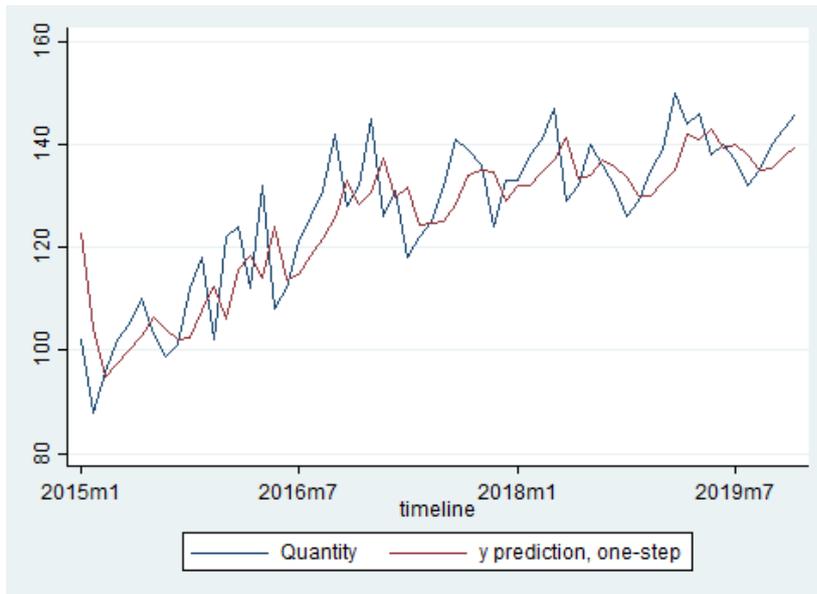


Fonte: Autores (2020)

### 3.2.11 Saída da Projeção Método ARMA(1,2)

Para esse modelo matemático, o erro MAPE é de 5,57% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método ARMA (1,2).

Figura 32 - Saída Gráfica do Método ARMA (1,2)

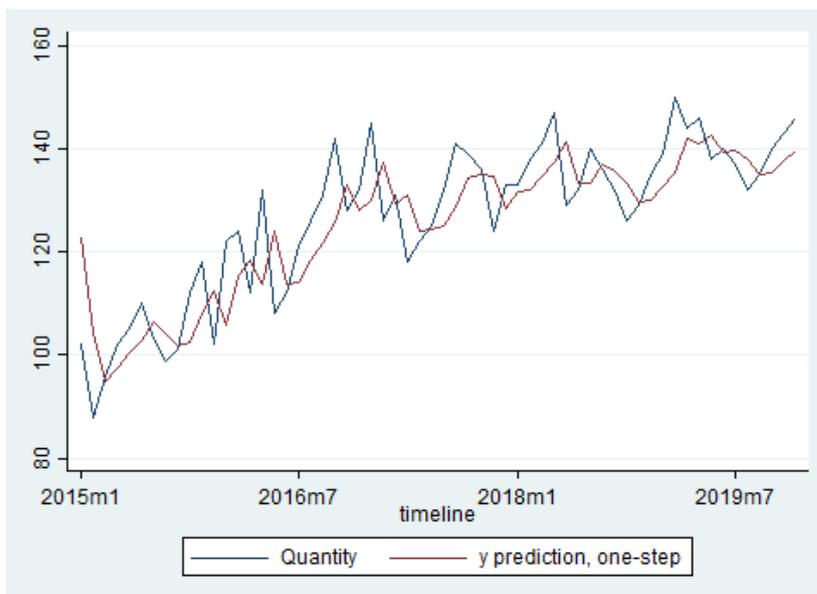


Fonte: Autores (2020)

### 3.2.12 Saída da Projeção Método ARMA(2,1)

Para esse modelo matemático, o erro MAPE é de 5,56% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método ARMA (2,1).

Figura 33 - Saída Gráfica do Método ARMA(2,1)

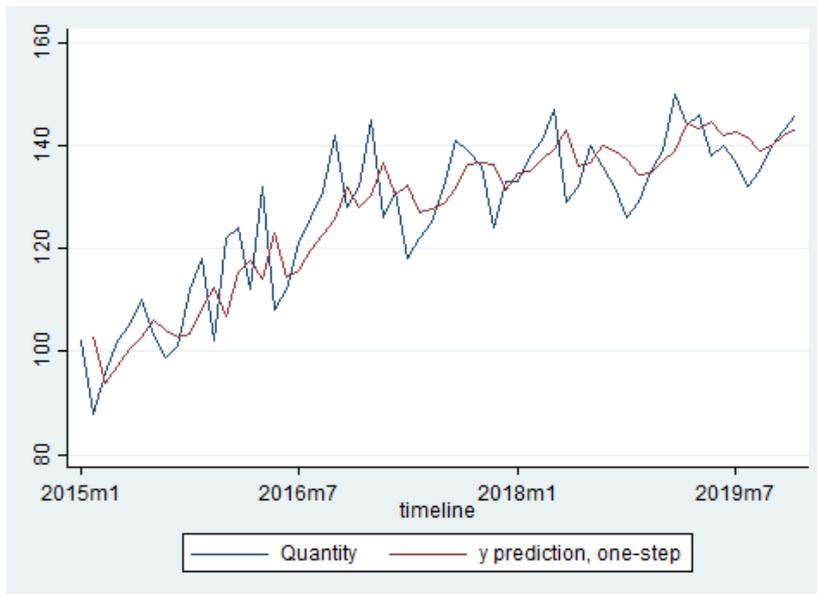


Fonte: Autores (2020)

### 3.2.13 Saída da Projeção Método ARMA(2,2)

Para esse modelo matemático, o erro MAPE é de 5,54% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o método ARMA (2,2).

Figura 34 - Saída Gráfica do Método ARMA (2,2)



Fonte: Autores (2020)

### 3.2.14 Saída da Projeção Método Redes Neurais

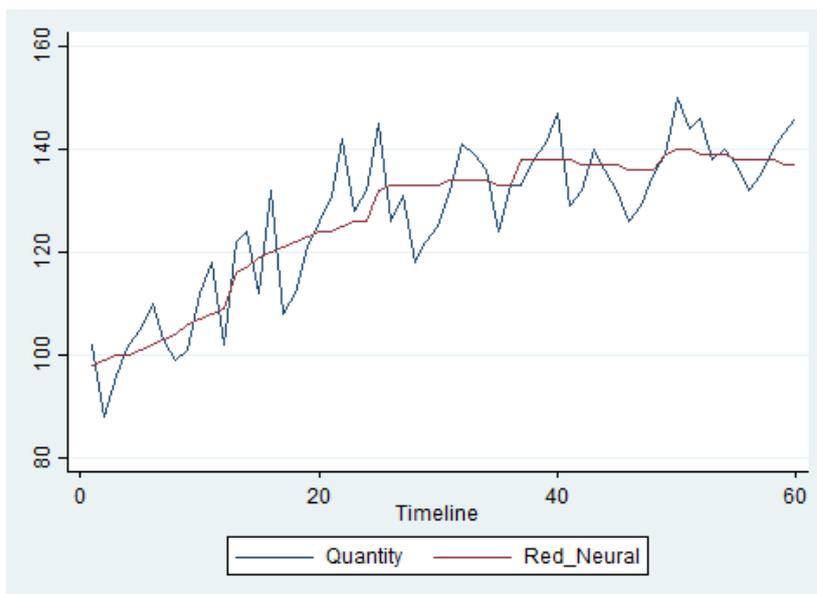
Para esse método de projeção foram realizados dois testes distintos, porém ambos com 75% do banco de dados sendo destinados para treino e 25% para testes da rede. A primeira rede considerou apenas uma camada oculta ao passo que na segunda rede foram consideradas duas. Todos os parâmetros utilizados para configurar as redes podem ser vistos na tabela a seguir e, posteriormente, é possível verificar a saída gráfica de cada uma delas:

Tabela 2 - Parâmetros das Redes Neurais para Projeção do Produto L

Parâmetros	Rede Neural 1	Rede Neural 2
% de treino	75	75
% de teste	25	25
Número de camadas	1	2
Número de neurônios (em cada camada oculta)	3	2
Função de ativação (Camada Oculta)	Tang. hiperbólica	Tang. hiperbólica
Função de ativação (Saída)	Identidade	Identidade
Erro quadrático (%)	5,561	4,422

Fonte: Autores (2020)

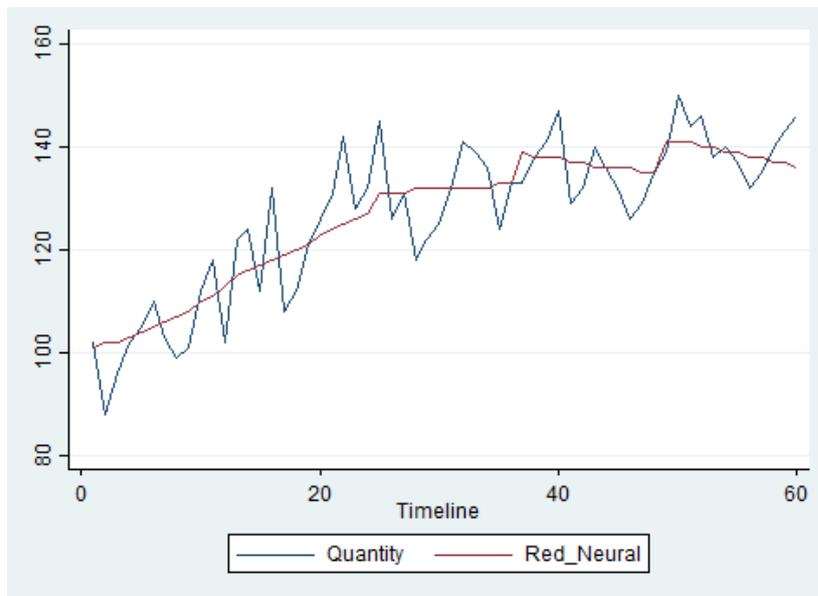
Figura 35 - Saída Gráfica do Método Rede Neural com 1 camada oculta



Fonte: Autores (2020)

Para esse modelo matemático, o erro MAPE é de 4,64% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o Perceptron de 1 camada oculta.

Figura 36 - Saída Gráfica do Método Rede Neural com 2 camadas ocultas



Fonte: Autores (2020)

Para esse modelo matemático, o erro MAPE é de 4,58% para a previsão dos últimos 12 períodos. Em azul destaca-se a demanda real do último período e em vermelho a projeção feita com o Perceptron de 2 camadas ocultas.

### 3.3 CONSOLIDAÇÃO DOS RESULTADOS OBTIDOS

Tabela 3 - Resultados obtidos através dos modelos de projeção de demanda

<b>Modelo</b>	<b>MAPE</b>	<b>MAPE %</b>
PERCEPTRON COM 2 CAMADAS OCULTAS	0,0458363	4,584
PERCEPTRON COM 1 CAMADA OCULTA	0,0464094	4,641
ARMA (2,2)	0,0553538	5,536
ARMA (2,1)	0,0556200	5,562
ARMA (1,2)	0,0557417	5,575
MÉDIA MÓVEL PONDERADA	0,0564330	5,644
ARMA (1,1)	0,0572202	5,723
MÉDIA MÓVEL SIMPLES	0,0577859	5,779
AR (1)	0,0579182	5,792
AR (2)	0,0580627	5,807
SUAUIZAÇÃO EXPONENCIAL DUPLA (MÉTOD0 DE BROWN)	0,0697018	6,971
MA (2)	0,0703953	7,040
AMORTECIMENTO EXPONENCIAL SIMPLES	0,0742558	7,426
MA (1)	0,0803697	8,037
SUAUIZAÇÃO EXPONENCIAL COM TENDÊNCIA (MÉTOD0 HOLT)	0,1167657	11,677

Fonte: Autores (2020)

#### 4. CONCLUSÃO

Objetivando estudar modelos de projeção de demanda desenvolvidos por teóricos, os conceitos por trás de algumas metodologias foram abordados e também aplicados a uma empresa de pequeno porte do setor de tecnologia, permitindo analisar a aplicabilidade ou não de cada método conforme as características peculiares do produto da empresa estudada. Para isso, através de métodos como média móvel ponderada, suavização exponencial dupla ou com tendência, amortecimento simples e redes neurais foi possível calcular o erro percentual médio absoluto (MAPE).

Diante de um vasto portfolio, foi necessária uma análise de Curva ABC para determinar o alcance a ser analisado e foi possível verificar que o “Produto L” representava maior parte do faturamento. A partir dos resultados obtidos, foram elencados os MAPE’s, correspondentes aos modelos utilizados, de forma crescente, com o intuito de avaliar quais os métodos mais assertivos para a previsão de demanda para o produto escolhido. A partir dos resultados, é possível verificar que, dos modelos testados, o modelo que gerou menor erro foi o Perceptron Multicamadas (2 camadas ocultas), ao passo que o menos adequado é o de Suavização Exponencial com Tendência (Método Holt), não sendo efetivo para este tipo de projeção.

Tendo em vista que atualmente a empresa não utiliza nenhum método de projeção de demanda, apenas a análise empírica de que se deve manter em estoque os produtos que são mais demandados pelos clientes, recomenda-se utilizar o método de Rede Neural Multicamadas (2 camadas ocultas). Apesar de existirem softwares pagos para a realização de projeção de demanda através dos métodos citados, há também uma gama de aplicativos que podem ser implementados a baixo custo. A implementação de um software dessa natureza auxiliaria a empresa a reverter a situação atual, onde há a ocupação desnecessária de espaço no estoque e capital imobilizado desnecessariamente.

## REFERÊNCIAS

- ASHTON, K. **That “Internet of Things” Thing**. RFID Journal, p. 4986, 2009.
- ASSOCIAÇÃO BRASILEIRA DE ENGENHARIA DE PRODUÇÃO (ABEPRO). Disponível em: <<http://portal.abepro.org.br/s>> Acesso em 09 de Outubro de 2019
- BACCI, A. B. **Combinação de métodos de séries temporais para previsão da demanda de café no Brasil**, 2007. Disponível em: [https://repositorio.unifei.edu.br/xmlui/bitstream/handle/123456789/1805/dissertacao\\_0031728.pdf?sequence=1&isAllowed=y](https://repositorio.unifei.edu.br/xmlui/bitstream/handle/123456789/1805/dissertacao_0031728.pdf?sequence=1&isAllowed=y). Acesso em 22 de Novembro 2019.
- BOURCHARDT, E. **Inteligência Artificial: Um Pouco da História e Avanços Atuais**. Disponível em: <<https://medium.com/@eliezerfb/intelig%C3%A9ncia-artificial-499fc2c4aa79>> Acesso em 9 de Novembro de 2019
- BRAGA, A. P., LUDERMIR, T. B. e CARVALHO, A. C. P. L. F. **Redes Neurais Artificiais: Teoria e aplicações**. Rio de Janeiro: LTC Editora, 2000.
- BRAGA, R. **Fundamentos e técnicas de administração financeira**. São Paulo: Atlas, 1989.
- BROWNLEE, J. **How to Create an ARIMA Model for Time Series Forecasting in Python**, 2017. Disponível em: <https://machinelearningmastery.com/arma-for-time-series-forecasting-with-python/>. Acesso em 16 de Novembro de 2019.
- BRYNJOLFSSON, E.; MCAFEE, A. **The second machine age: Work, progress, and prosperity in a time of brilliant technologies**. WW Norton & Company: 2014.
- CALDEIRA, A. M., et al. **Inteligência Computacional Aplicada à Administração, Economia e Engenharia em Matlab**. São Paulo: Thomson Learning, 2007.
- CARVALHO, J. M. C. **Logística**. 3. ed. Lisboa: Edições Silabo, 2002
- CETAX. Data Science: **O Que É, Conceito e Definição**. Disponível em: <<https://www.cetax.com.br/blog/data-science/>> Acesso em 03 de Novembro de 2019
- COBRA, M. **Administração de marketing**. São Paulo: Cobra, 2005.
- COMPUTER WORLD. **Por Que Python é A Melhor Linguagem de Programação Para Iniciantes**. Disponível em: <<https://computerworld.com.br/2018/10/05/por-que-python-e-a-melhor-linguagem-de-programacao-para-iniciantes/>> Acesso em 03 de Novembro de 2019
- CORRÊA, et al.– **Planejamento, Programação e Controle da Produção**. 4ª ed. São Paulo: Atlas, 2001.
- COSTA, C. L. **Internet das Coisas: Um Estudo Exploratório em Agronegócios**. Disponível em: <<https://www.ufrgs.br/cienagro/wp-content/uploads/2018/10/Internet->

das-coisas-IOT-um-estudo-explorat%C3%B3rio-em-agroneg%C3%B3cios.docx-Cain%C3%A3-Lima-Costa.pdf> Acesso em 17 de Novembro de 2019

COSTA, R. **Conceitos Básicos de Redes Neurais Artificiais**. Disponível em: <<https://www.linkedin.com/pulse/conceitos-b%C3%A1sicos-de-redes-neurais-artificiais-ruan-costa>> Acesso em 23 de Novembro de 2019

DAVIES, N. **A Europa em Guerra (1939-1945)**. Lisboa: Edição 70, 2008  
DAVIS, M.; AQUILANO, N.; CHASE, R. **Fundamentos da Administração da Produção**. Porto Alegre: Bookman, 2001.

DATA SCIENCE ACADEMY. **Deep Learning Book**. Disponível em: <<http://deeplearningbook.com.br/>> Acesso em 22 de Fevereiro de 2020

DATA SCIENCE ACADEMY. **Inteligência Artificial Não Está Eliminando Empregos; Está Eliminando Modelos de Negócio**. Disponível em: <<http://datascienceacademy.com.br/blog/inteligencia-artificial-nao-esta-eliminando-empregos-esta-eliminando-modelos-de-negocio/>> Acesso em 03 de Novembro de 2019

DIAS, M. A. P. **Administração de materiais: Princípios, conceitos e gestão**. 5. ed. São Paulo: Atlas, 2005.

EINSTEIN, A. **25 Frases de Albert Einstein**. 2020. Disponível em: <<https://escolaeducacao.com.br/frases-de-albert-einstein/>> Acesso em 08 de Junho de 2020

FACURE, M. **Funções de Ativação** Disponível em: <<https://matheusfacure.github.io/2017/07/12/activ-func/>> Acesso em 22 de Novembro de 2019

FERNANDES, C. **Máquina Enigma**. Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/historiag/maquina-enigma.htm>. Acesso em 09 de Novembro de 2019.

FERNANDES, F. C. F.; GODINHO, M. **Planejamento e Controle da Produção: Dos Fundamentos ao Essencial**. 1ª ed. São Paulo: Atlas, 2017.

FERREIRA, F. **Os 23 Princípios de Asilomar**. Disponível em: <<https://www.deviant.com.br/noticias/ciencia/os-23-principios-de-asilomar/>> Acesso em 03 de Novembro de 2019

FINOCCHIO, M. A. F. **Noções de Redes Neurais Artificiais**. Paraná: UTFP, 2014

FIRJAN. **Indústria 4.0: A Internet das Coisas**. Rio de Janeiro: 2016 Disponível em: <<https://www.firjan.com.br>> Acesso em 18 de Outubro de 2019

FORBES. **Why 2017 Is The Year Of Artificial Intelligence**. Disponível em: <<https://www.forbes.com/sites/forbestechcouncil/2017/02/27/why-2017-is-the-year-of-artificial-intelligence/#6b9874a757a1>> Acesso em 09 de Novembro de 2019

FUTURE OF LIFE INSTITUTE. **Asilomar AI Principles**. Disponível em: <<https://futureoflife.org/ai-principles/?cn-reloaded=1>> Acesso em 03 de Novembro de 2019

GARTNER. **Top Trends On The Gartner Hype Cycle For Artificial Intelligence**. Disponível em: <<https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/>> Acesso em 03 de Novembro de 2019

GRANATYR, J. **Teste de Turing**. Disponível em: <<https://iaexpert.com.br/index.php/2016/07/19/historico-da-ia-teste-de-turing/>> Acesso em 09 de Novembro de 2019

GUIMARÃES, A. M. C. **Empresas de Gestão Conservadora: Potencial da Previsão de Demanda e Simulação Computacional**. Rio de Janeiro, 2008

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2ed. New Jersey: Prentice-Hall Inc., 1999.

HELLINGER, A.; SEEGER, H. **Cyber-Physical Systems - Driving force for innovation in mobility, health, energy and production**. National Academy of Science and Engineering: 2011

HERMANN, M.; PENTEK, T.; OTTO, B. **Design principles for industrie 4.0 scenarios**. Proceedings of the Annual Hawaii International Conference on System Sciences: 2016.

KANG, H. S., J. Y. LEE, S. CHOI, H. KIM, J. H. PARK, J.Y. SON, B. H. KIM. S. D. NOH **Smart manufacturing: Pastresearch, present findings, and future directions**. International Journal of Precision Engineering and Manufacturing-Green Technology: 2016

KHAN ACADEMY. **Neuron Biology: An Overview of Neuron Structure and Function**. Disponível em: <<https://pt.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/overview-of-neuron-structure-and-function>> Acesso em 22 de Fevereiro de 2020

KOVÁCS, Z. L. **Redes Neurais Artificiais Fundamentos e Aplicações**. São Paulo: Editora Livraria da Física, 2002.

LANEY, D. **3D Data Management: Controlling Data Volume, Velocity and Variety**. Disponível em: < <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>> Acesso em 10 de Novembro de 2019

MANCUZO, F. **Análise e Previsão de Demanda: Estudo de caso em uma empresa distribuidora de rolamentos**. Porto Alegre, 2003. Disponível em: <<http://www.producao.ufrgs.br/arquivos/publicacoes/Fernando%20Mancuzo.pdf>> Acesso em 09 de Outubro de 2019

MARMELLO, D. **A CURVA ABC PARA SUPRIMENTOS INDIRETOS**. Disponível em: <<https://mundo.gimba.com.br/supply-chain/a-curva-abc-para-suprimentos-indiretos/>> Acesso em 31 de março de 2020

MARTIN, J. **Amazon Go: O Futuro Está Próximo**. Disponível em: <<https://www.workana.com/blog/pt/emprendimentopt/amazon-go-o-futuro-esta-chegando/>> Acesso em 22 de Fevereiro de 2020

MARTINS, F. A. S. **Técnicas de Previsão**. Disponível em: <<https://sistemas.eel.usp.br/docentes/arquivos/5840917/LOQ4001/Cap3.TecnicasdePrevisoes.pdf>> Acesso em 20 de Outubro de 2019

MARTINS, L. **Tecido Nervoso**. Disponível em: <<https://www.infoescola.com/biologia/tecido-nervoso/>> Acesso em 24 de Novembro de 2019

MARTINS, P. G; LAUGENI, F. P. **Administração da Produção**. 3ed. São Paulo: Saraiva, 2015

MAKRIDAKIS, S. **Metaforecasting: Ways of Improving Forecasting Accuracy and Usefulness**. France, 1988. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0169207088901124>> Acesso em 19 de Outubro de 2019

MCKENDRICK, J. **Artificial Intelligence Isn't Killing Jobs. It's Killing Business Models**. Disponível em: <<https://www.forbes.com/sites/joemckendrick/2018/01/25/artificial-intelligence-isnt-killing-jobs-its-killing-business-models/#1297e8065ea0>> Acesso em 03 de Novembro de 2019

MITIUYE, E. Y; SILVA, M. N. M; PEREIRA, M. **Curva ABC: Princípios e Aplicabilidade Empresarial**. Disponível em: <[http://faef.revista.inf.br/imagens\\_arquivos/arquivos\\_destaque/3rMaREuEuoWdCf2\\_2013-4-23-17-31-0.pdf](http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/3rMaREuEuoWdCf2_2013-4-23-17-31-0.pdf)> Acesso em 31 de Março de 2020

MONKS, J. G. **Administração da Produção**. São Paulo: McGraw-Hill, 1987

MOREIRA, D. A. **Administração da produção e operações**. 2. ed. São Paulo: Cengage Learning, 2008

MOREIRA, S. **Rede Neural Perceptron Multicamadas**. Disponível em: <<https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>> Acesso em 20 de Novembro de 2019

MORETTIN, P. A.; TOLOI, C. M. C. **Previsão de séries temporais**. 2. ed. São Paulo: Atual, 2006.

NARASIMHAN, S.; MCLEAVEY, D. W.; BILLINGTON, P. **Production Planning and Inventory Control**. 2ed. New Jersey: Prentice Hall, 1995.

NASCIMENTO, R. **Afinal, O Que É Big Data?**. Disponível em: <<http://marketingpordados.com/analise-de-dados/o-que-e-big-data-%F0%9F%A4%96/>> Acesso em 10 de Novembro de 2019

NIELSEN, M. **Neural Networks And Deep Learning**. Disponível em: <<http://neuralnetworksanddeeplearning.com/chap1.html>> Acesso em 22 de Fevereiro de 2020

OLIVEIRA, C. S. et. al. **As Redes Neurais de Hopfield e Multi-Layer Perceptrons Formando uma Arquitetura Neural Híbrida (MLP+H) com Características Próprias**. Disponível em: <[http://abricom.org.br/wp-content/uploads/2016/09/6CBRN\\_005.PDF](http://abricom.org.br/wp-content/uploads/2016/09/6CBRN_005.PDF)> Acesso em 21 de Novembro de 2019

PELLEGRINI, F. R. **Metodologia para Implementação de Sistemas de Previsão de Demanda**. Disponível em: <<http://www.producao.ufrgs.br/arquivos/publicacoes/Fernando%20R%20Pellegrini.pdf>> Acesso em 15 de Outubro de 2019.

PEREIRA, S. A. **Estudo comparativo entre modelos de previsão de demanda: ensaio em um produto classe a de uma empresa de perfumes e cosméticos**. Ceará: ENEGEP. 2006

PESSOA, K; NOBRE, D.; VERAS, B. **Análise Comparativa Entre os Métodos de Previsão de Demanda em uma Indústria de Higiene Pessoal, Perfumaria e Cosméticos Localizada em Teresina**. Alagoas, 2018. Disponível em: <[https://www.researchgate.net/publication/329069826\\_Analise\\_comparativa\\_entre\\_os\\_metodos\\_de\\_previsao\\_de\\_demanda\\_em\\_uma\\_industria\\_de\\_higiene\\_pessoal\\_perfumaria\\_e\\_cosmeticos\\_localizada\\_em\\_Teresina-PI](https://www.researchgate.net/publication/329069826_Analise_comparativa_entre_os_metodos_de_previsao_de_demanda_em_uma_industria_de_higiene_pessoal_perfumaria_e_cosmeticos_localizada_em_Teresina-PI)> Acesso em 09 de Outubro de 2019

PINSODORF, M. **5 Orientações Éticas Sobre IA Que Todo Executivo C-Level Deve Ler**. Disponível em: <<https://cio.com.br/5-orientacoes-eticas-sobre-ia-que-todo-executivo-c-level-deve-ler/>> Acesso em 03 de Novembro de 2019

PORTAL EDUCAÇÃO. **O Que É Conferência de Asilomar?** Disponível em: <<https://www.portaleducacao.com.br/conteudo/artigos/enfermagem/o-que-e-conferencia-de-asilomar/40319>> Acesso em 03 de Novembro de 2019

PINDYCK, R. S., RUBINFELD, D. L. **Econometric Models & Economic Forecasts**. 8 ed. Singapore: McGraw Hill Book Co., 2013.

PWC. **Indústria 4.0: Digitalização Como Vantagem Competitiva no Brasil**. Disponível em: <<https://www.pwc.com.br/pt/publicacoes/servicos/assets/consultoria-negocios/2016/pwc-industry-4-survey-16.pdf>> Acesso em 10 de Novembro de 2019

REIS, B. **Redes Neurais – Funções de Ativação**. Disponível em: <<http://www.decom.ufop.br/imobilis/redes-neurais-funcoes-de-ativacao/>> Acesso em 22 de Novembro de 2019

ROMANELLI, B. **Clusterização de dados de vibração na perfuração de poços de petróleo através de redes neurais não supervisionadas**. Disponível em: <[https://www.maxwell.vrac.puc-rio.br/25049/25049\\_6.PDF](https://www.maxwell.vrac.puc-rio.br/25049/25049_6.PDF)> Acesso em 22 de Fevereiro de 2020

RUSSWURM, S. **Industry 4.0 - from vision to reality**. Background Information: 2014.

SANTOS, B. P. et al. **Internet das Coisas: Da Teoria à Prática**. Disponível em: <<https://homepages.dcc.ufmg.br/~mmvieira/cc/papers/internet-das-coisas.pdf>> Acesso em 10 de Novembro de 2019

SHARMA, A. **Restricted Boltzmann Machines – Simplified**. Disponível em: <<https://towardsdatascience.com/restricted-boltzmann-machines-simplified-eab1e5878976>> Acesso em 21 de Novembro de 2019

SHARMA, A. **Understanding Activation Functions in Neural Networks**. Disponível em: <<https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>> Acesso em 21 de Novembro de 2019

SOKOLOVSKY, O. **Managing Business Leadership Expectations Around The AI Hype**. Disponível em: <[https://www.dataversity.net/managing-business-leadership-expectations-around-ai-hype/?fbclid=IwAR0hoJLaSz9kSnIH5InIxPECI1I3T7IEZwQymEn4uB2fyHDgPzXsnwI\\_\\_ns](https://www.dataversity.net/managing-business-leadership-expectations-around-ai-hype/?fbclid=IwAR0hoJLaSz9kSnIH5InIxPECI1I3T7IEZwQymEn4uB2fyHDgPzXsnwI__ns)> Acesso em 03 de Novembro de 2019

SOUSA, F. B. **Análise de Modelo de Hopfield com Topologia de Rede Complexa**. Disponível em: <[https://teses.usp.br/teses/disponiveis/55/55134/tde-30012014-111520/publico/Sousa2013\\_masterthesis.pdf](https://teses.usp.br/teses/disponiveis/55/55134/tde-30012014-111520/publico/Sousa2013_masterthesis.pdf)> Acesso em 21 de Novembro de 2019

STOCK, T., SELIGER, G. **Opportunities of Sustainable Manufacturing in Industry 4.0**. 2016.

SCHUH, B., POTENTE, T., WESCH-POTENTE, C., WEBER, A., PROTE, J., **Collaboration Mechanisms to Increase Productivity in the Context of Industrie 4.0**. 2015.

SILVA, J. M. N. **Redes Neurais Artificiais: Rede Hopfield e Redes Estocásticas**. Disponível em: <[http://www.ic.uff.br/~jsilva/monografia\\_RNA.pdf](http://www.ic.uff.br/~jsilva/monografia_RNA.pdf)> Acesso em 20 de Novembro de 2019

TATIBANA, C. Y.; KAETSU, D. Y. **Uma Introdução Às Redes Neurais**. Disponível em: <<https://ceticismo.net/ciencia-tecnologia/uma-introducao-as-redes-neurais/>> Acesso em 23 de Novembro de 2019

THANO. **Máquinas de Boltzmann Restritas**. Disponível em: <<https://matheusfacure.github.io/2017/07/30/RBM/>> Acesso em 21 de Novembro de 2019

THOMAS, C.; CHANG, J.; BAAN, W. **How Advanced Industrial Companies Should Approach Artificial Intelligence Strategy**. Disponível em: <<https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-advanced-industrial-companies-should-approach-artificial-intelligence-strategy>> Acesso em 03 de Novembro de 2019

TUBINO, D. F. **Manual de Planejamento e Controle da Produção**. 2ª ed. São Paulo: Atlas, 2000.

VDE-DKE. **The German Standardization Roadmap Industrie 4.0**. Disponível em: <<https://www.din.de/blob/65354/57218767bd6da1927b181b9f2a0d5b39/roadmap-i4-0-e-data.pdf>> Acesso em 22 de Outubro de 2019

VEZA, I., M. MLADINEO, N. GJELDUM **Managing Innovative Production Network of Smart Factories**. 2015.

VILELA A. L. M. **Anatomia e Fisiologia Humana – Coordenação e Regulação**. Disponível em: <<http://www.afh.bio.br/basicos/Nervoso1.htm>> Acesso em 24 de Novembro de 2019

WERNER, L.; RIBEIRO, J. L. D. **Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais**, 2003. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-530X2003000100005](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2003000100005). Acesso em 16 de Novembro de 2019.

WITTEK, P. **Hopfield Network**. Disponível em: <<https://www.sciencedirect.com/topics/computer-science/hopfield-network>> Acesso em 22 de Fevereiro de 2020

XIA, F. et al. **Internet of Things**. Int. J. Commun. Syst. 2012.

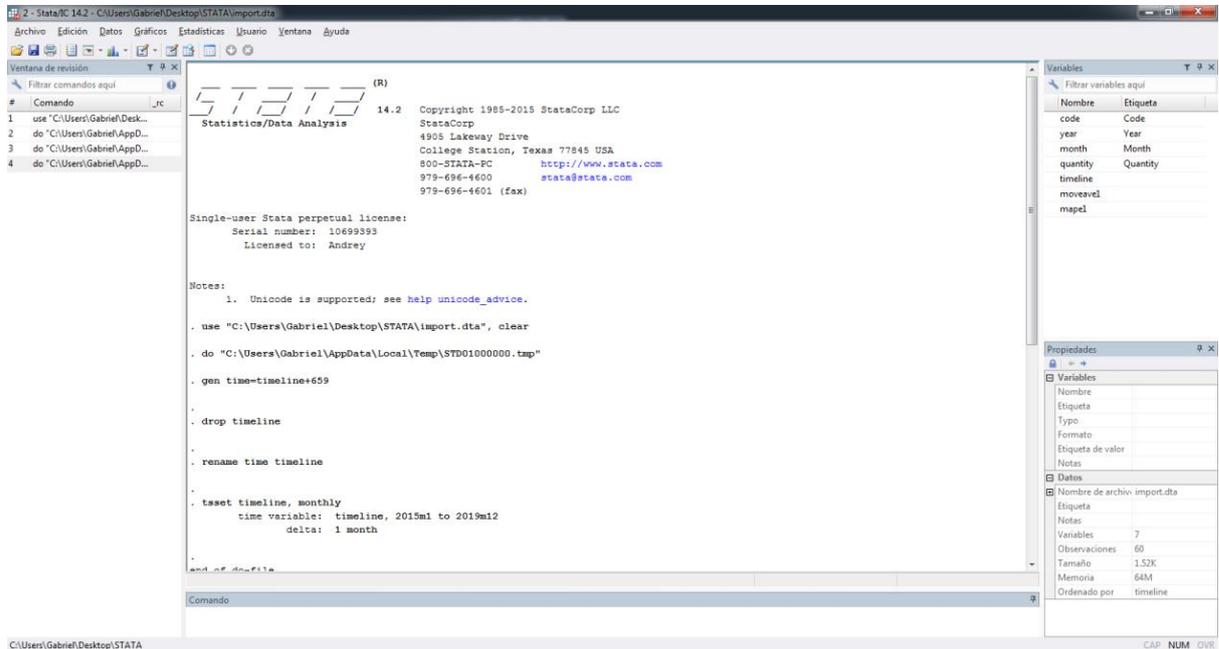
YUGE, D. **Alan Turing: O Pai da Computação**. Disponível em: <<https://www.tecmundo.com.br/ciencia/142291-ha-65-anos-morria-alan-turing-pai-computacao-da-ia.htm>> Acesso em 16 de Novembro de 2019

ZHOU, K., LIU, T., ZHOU, L. **Industry 4.0: Towards future industrial opportunities and challenges**. 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, 2015

ZONER, T.; BONGIOVANNI, S. **Gestão de resíduos sólidos do município de Tibagi: Desempenho da Coleta Seletiva**. Disponível em: <<https://repositorio.unesp.br/bitstream/handle/11449/126956/ISSN2236-1170-2012-08-08-1591-1601.pdf?sequence=1>> Acesso em 20 de Outubro de 2019

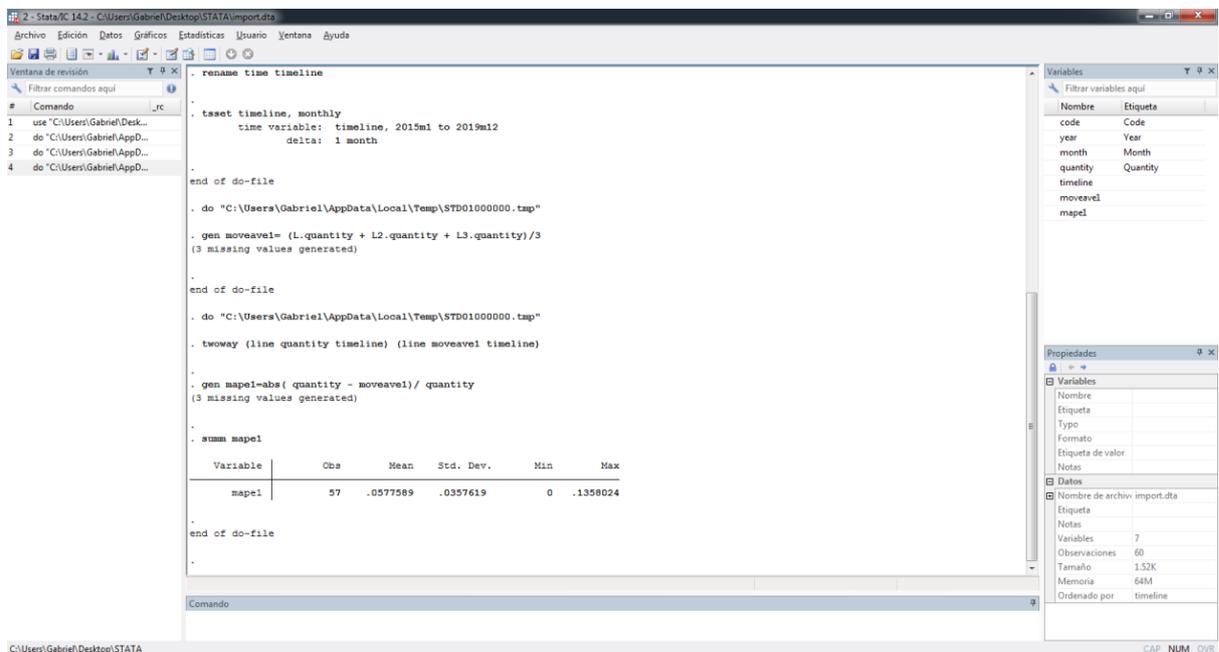
## ANEXO A – TELA DO SOFTWARE STATA

Figura 37 - Tela com parametrização do algoritmo



Fonte: Autores (2020)

Figura 38 - Tela do Software Stata com o algoritmo da Média Móvel



Fonte: Autores (2020)

## ANEXO B – SAÍDA DO SOFTWARE SSPS PARA PERCEPTRON COM 1 CAMADA OCULTA

\*Multilayer Perceptron Network.

MLP Quantity (MLEVEL=S) WITH Year Month Timeline

/RESCALE COVARIATE=STANDARDIZED

/PARTITION TRAINING=75 TESTING=25 HOLDOUT=0

/ARCHITECTURE AUTOMATIC=YES (MINUNITS=1 MAXUNITS=50)

/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE

LAMBDAINITIAL=0.0000005

SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5

MEMSIZE=1000

/PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION

/PLOT NETWORK

/SAVE PREDVAL

/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON

(MAXTIME=15) MAXEPOCHS=AUTO

ERRORCHANGE=1.0E-4 ERRORRATIO=0.001

/MISSING USERMISSING=EXCLUDE .

### Observações

Saída criada		07-JUN-2020 00:48:31
Comentários		
Entrada	Conjunto de dados ativo	ConjuntodeDados1
	Filtro	<none>
	Ponderação	<none>
	Arquivo Dividido	<none>
	Nº de linhas em arquivo de dados de trabalho	60
Tratamento de valor omissos	Definição de omissos	Os valores omissos para o usuário e do sistema são tratados como omissos.

Casos utilizados		As estatísticas têm como base os casos com dados válidos para todas as variáveis usadas pelo procedimento.
Tratamento de ponderação		Não se aplica
Recursos	Tempo do processador	00:00:01,84
	Tempo decorrido	00:00:01,16
Variáveis Criadas ou Modificadas	Valor previsto	MLP_PredictedValue

Fonte: SSPS (2020)

### Resumo de processamento de casos

		N	Porcentagem
Amostra	Treinamento	44	73,3%
	Testes	16	26,7%
Válido		60	100,0%
Excluídos		0	
Total		60	

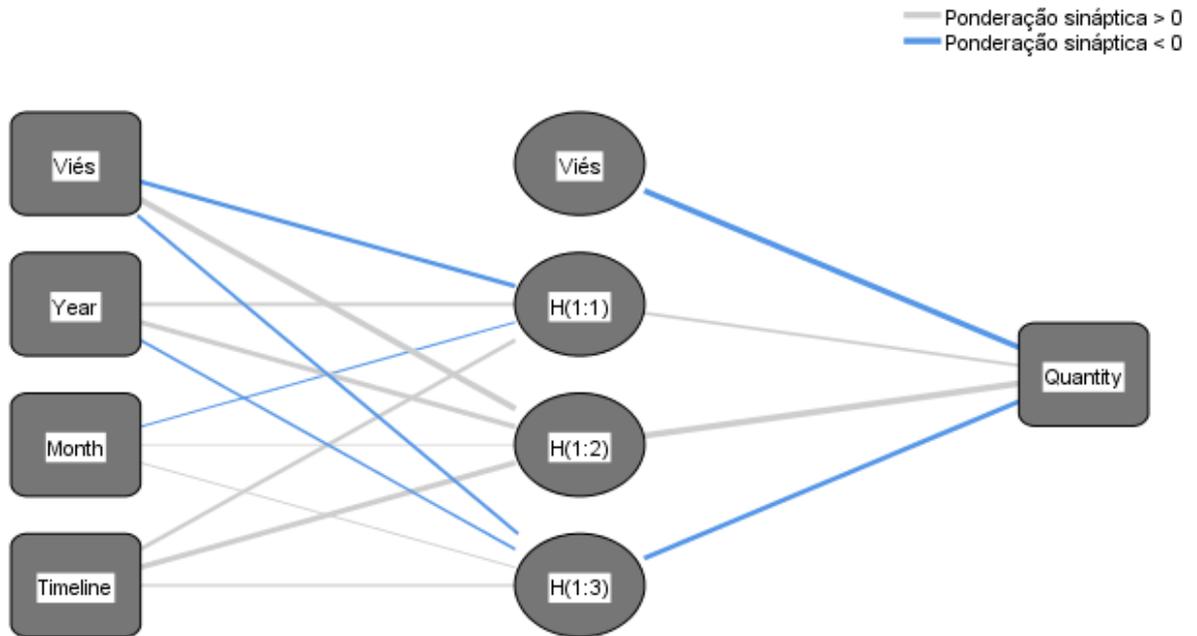
Fonte: SSPS (2020)

### Informações de rede

Camada de entrada	Covariáveis	1	Year
		2	Month
		3	Timeline
	Número de unidades <sup>a</sup>	3	
Método de reescalonamento para covariáveis		Padronizado	
Camadas ocultas	Número de camadas ocultas		1
	Número de Unidades na Camada Oculta 1 <sup>a</sup>		3
	Função de ativação		Tangente hiperbólica
Camada de saída	Variáveis dependentes	1	Quantity
	Número de unidades		1
	Método de reescalonamento para dependentes de escala		Padronizado
	Função de ativação		Identidade

Fonte: SSPS (2020)

Figura 38 - Diagrama da Rede com 1 Camada Oculta Elaborada para o Estudo de Caso



Função de ativação de camada oculta: Tangente hiperbólica

Função de ativação de camada de saída: Identidade

Fonte: SSPS (2020) adaptado pelos Autores (2020)

### Resumo do modelo

Treinamento	Soma dos erros	5,561
	Erro relativo	,259
	Regra de parada usada	1 etapas consecutivas sem diminuição de erros <sup>a</sup>
	Tempo de treinamento	0:00:00,00
Testes	Soma dos erros	1,564
	Erro relativo	,153

Fonte: SSPS (2020)

## ANEXO C – SAÍDA DO SOFTWARE SSPS PARA PERCEPTRON COM 2 CAMADAS OCULTAS

\*Multilayer Perceptron Network.

```
MLP Quantity (MLEVEL=S) WITH Month Timeline Year
/RESCALE COVARIATE=STANDARDIZED DEPENDENT=STANDARDIZED
/PARTITION TRAINING=75 TESTING=25 HOLDOUT=0
/ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=2 (NUMUNITS=AUTO)
HIDDENFUNCTION=TANH
  OUTPUTFUNCTION=IDENTITY
/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE
LAMBDAINITIAL=0.0000005
  SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5
MEMSIZE=1000
/PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION
/PLOT NETWORK
/SAVE PREDVAL
/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON
(MAXTIME=15) MAXEPOCHS=AUTO
  ERRORCHANGE=1.0E-4 ERRORRATIO=0.001
/MISSING USERMISSING=EXCLUDE .
```

### Resumo de processamento de casos

	N	Porcentagem
Amostra		
Treinamento	45	75,0%
Testes	15	25,0%
Válido	60	100,0%
Excluídos	0	
Total	60	

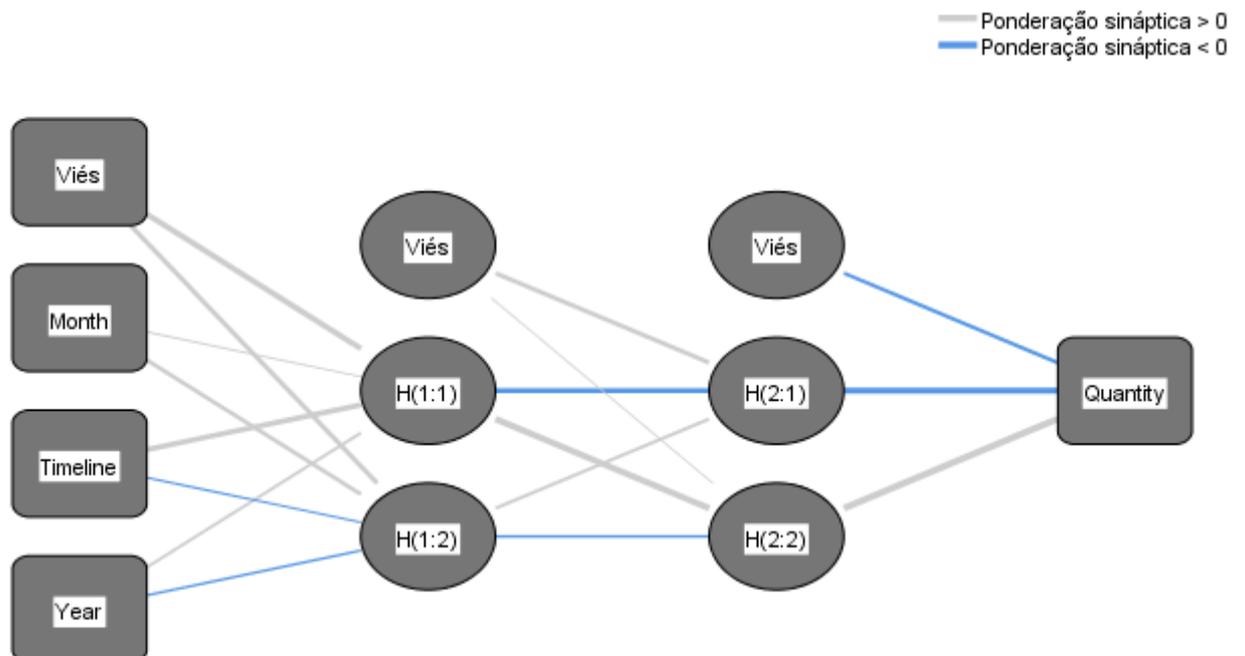
Fonte: SSPS (2020)

### Informações de rede

Camada de entrada	Covariáveis	1	Month
		2	Timeline
		3	Year
	Número de unidades <sup>a</sup>		3
Método de reescalonamento para covariáveis		Padronizado	
Camadas ocultas	Número de camadas ocultas		2
	Número de Unidades na Camada Oculta 1 <sup>a</sup>		2
	Número de Unidades na Camada Oculta 2 <sup>a</sup>		2
	Função de ativação		Tangente hiperbólica
Camada de saída	Variáveis dependentes	1	Quantity
	Número de unidades		1
	Método de reescalonamento para dependentes de escala		Padronizado
	Função de ativação		Identidade

Fonte: SSPS (2020)

Figura 39 - Diagrama da Rede com 2 Camadas Ocultas Elaborada para o Estudo de Caso



Função de ativação de camada oculta: Tangente hiperbólica

Função de ativação de camada de saída: Identidade

Fonte: SSPS (2020) adaptado pelos Autores (2020)

**Resumo do modelo**

Treinamento	Soma dos erros quadráticos	4,422
	Erro relativo	,201
	Regra de parada usada	1 etapas consecutivas sem diminuição de erros <sup>a</sup>
	Tempo de treinamento	0:00:00,00
Testes	Soma dos erros	1,832
	Erro relativo	,298

Fonte: SSPS (2020)