

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS**  
**PROGRAMA DE MESTRADO EM ENGENHARIA DE PRODUÇÃO E SISTEMAS**

**PREVISÃO DE DEMANDA ATRAVÉS DE REDES NEURAIAS**  
**EM UM AMBIENTE DE OMNICANALIDADE NO VAREJO**

**Eduardo Correia Lima Linhares**

**2021**

**PREVISÃO DE DEMANDA ATRAVÉS DE REDES NEURAS EM UM  
AMBIENTE DE OMNICANALIDADE NO VAREJO**

Eduardo Correia Lima Linhares

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador: Ricardo Luiz Machado, *Dr.*

Goiânia  
Setembro, 2021

L755p Linhares, Eduardo Correia Lima  
Previsão de demanda através de redes neurais em um  
ambiente de omnicanalidade no varejo / Eduardo Correia  
Lima Linhares.-- 2021.  
92 f.; il.;

Texto em português com resumo em inglês  
Dissertação (mestrado) -- Pontifícia Universidade  
Católica de Goiás, Escola de Engenharia, Goiânia,  
2021

Inclui referências, f. 85-92

1. Logística empresarial. 2. Oferta e procura. 3.  
Projeção. I.Machado, Ricardo Luiz. II.Pontifícia Universidade  
Católica de Goiás - Programa de Pós-Graduação em Engenharia  
de Produção e Sistemas - 2021. III. Título.

CDU: Ed. 2007 -- 658.7(043)

# PREVISÃO DE DEMANDA ATRAVÉS DE REDES NEURAIS EM UM AMBIENTE DE OMNICANALIDADE NO VAREJO

Eduardo Correia Lima Linhares

Esta Dissertação julgada adequada para obtenção do título de Mestre em Engenharia de Produção e Sistemas e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás em setembro de 2021.



---

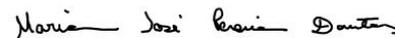
Prof. Marcos Lajovic Carneiro Dr.  
Coordenador do Programa de Pós-Graduação em  
Engenharia de Produção e Sistemas

## Banca Examinadora:



---

Prof. Ricardo Luiz Machado, Dr.  
Orientador



---

Prof.ª Maria José Pereira Dantas, Dr.ª  
Examinadora Interna - PUCGO



---

Prof.ª Eliana Tiba Gomes Grande, Dr.ª  
Examinadora Externa - IFG

## RESUMO

Dado o desenvolvimento do varejo rumo ao ambiente de omnicanalidade e o consequente aumento de fontes de informações disponíveis, a previsão de demanda se torna cada vez mais complexa e a busca por soluções que se adaptam a este cenário se torna relevante. Através de uma revisão bibliográfica, foram analisados os principais modelos de previsão de demanda que utilizam algoritmos de inteligência artificial no período de 2010 a 2020, assim como seus resultados e comparações com modelos tradicionais. Este trabalho apresenta um modelo de análise de dados que adota a combinação de dois algoritmos de *machine learning* (*hierarchical clustering* e redes neurais artificiais) para previsão de demanda a partir do treinamento em dados reais de um período de dois anos de um varejo local em um ambiente de omnicanalidade contendo atributos internos e externos. O modelo demonstrou, através dos indicadores de previsibilidade MAPE, MAE e RMSE, valores dentro do alcance dos resultados obtidos pela literatura para previsão de demanda, mesmo em um ambiente de omnicanalidade. As percepções deste estudo podem ser utilizadas para definição das políticas de ressuprimento e alocação de produtos nos diferentes canais de atendimento. A generalização deste modelo tem oportunidade para pesquisas futuras, levando em consideração outros atributos relevantes para a previsão de demanda como o perfil do consumidor, assim como variações do algoritmo para torná-lo mais robusto.

**Palavras-chave:** cadeia de suprimentos; *phygital*; *machine learning*; projeção de demanda

## ABSTRACT

Given the retail development towards an omnichannel environment and the consequent increase in available information sources, demand forecasting becomes increasingly complex. Thereby, the search for solutions that adapt to this scenario becomes relevant. A literature review analyzed the main demand prediction models using artificial intelligence algorithms from 2010 to 2020, as well as their results and comparisons with traditional models. This work presents a data analysis model that adopts the combination of two machine learning algorithms (hierarchical clustering and artificial neural networks) to forecast demand from training on two-year data from a local retailer in an omnichannel environment containing internal and external attributes. The model demonstrated, by the accuracy measures MAPE, MAE, and RMSE, values consistent with the results obtained in the literature for demand forecasting, despite the omnichannel environment. The insights of this research can be applied in fulfillment and allocation policies in different channels. Therefore, the generalization of this model provides the opportunity for further research, taking into account other attributes, such as the consumer profile, as well as variations of the algorithm to improve its robustness.

**Key-words:** supply chain; phygital; machine learning; demand projection

# SUMÁRIO

<b>LISTA DE ABREVIATURAS .....</b>	<b>7</b>
<b>LISTA DE TABELAS.....</b>	<b>10</b>
<b>LISTA DE QUADROS .....</b>	<b>11</b>
<b>LISTA DE FIGURAS .....</b>	<b>12</b>
<b>1 INTRODUÇÃO .....</b>	<b>13</b>
1.1 CONTEXTO .....	13
1.2 PROBLEMÁTICA .....	17
1.2.1 Questão geral.....	18
1.2.2 Questões específicas .....	19
1.3 OBJETIVOS DA PESQUISA .....	19
1.3.1 Objetivo geral.....	20
1.3.2 Objetivos específicos .....	20
1.4 JUSTIFICATIVA .....	21
1.5 SÍNTESE METODOLÓGICA .....	22
1.6 ESTRUTURA DO TRABALHO .....	22
<b>2 REVISÃO DA LITERATURA .....</b>	<b>24</b>
2.1 OMNISCANALIDADE.....	25
2.2 <i>MACHINE LEARNING</i> E REDES NEURAS ARTIFICIAIS .....	27
2.3 MÉTRICAS DE DESEMPENHO .....	34
2.3.1 MSE.....	36
2.3.2 RMSE.....	36
2.3.3 MAPE.....	36

<b>2.3.4 MAE</b> .....	37
<b>2.4 METODOLOGIAS PARA PREVISÃO DE DEMANDA NA LITERATURA</b> .....	38
2.4.1 ARTIGOS COM ATRIBUTOS EXCLUSIVAMENTE INTERNOS .....	39
2.4.2 ARTIGOS COM ATRIBUTOS INTERNOS E EXTERNOS .....	45
2.5 ANÁLISE DE DADOS .....	57
<b>3 ABORDAGEM METODOLÓGICA</b> .....	59
3.1 DELIMITAÇÃO DA PESQUISA .....	60
3.2 OBJETO DE PESQUISA .....	60
3.3 MÉTODO DE PESQUISA .....	61
3.4 INSTRUMENTOS DA PESQUISA .....	66
<b>4 RESULTADOS</b> .....	67
4.1 CLUSTERIZAÇÃO .....	68
4.2 REDES NEURAIS ARTIFICIAIS .....	70
4.3 COMPARATIVO DAS MÉTRICAS DE DESEMPENHO .....	73
4.4 IMPLICAÇÕES GERENCIAIS .....	79
<b>5 CONSIDERAÇÕES FINAIS</b> .....	81
<b>REFERÊNCIAS</b> .....	83

# LISTA DE ABREVIATURAS

ADAM	<i>Adaptative Moment Estimation</i>
AHN	<i>Artificial Hydrocarbon Networks</i>
ANFIS	<i>Adaptive Network-based Fuzzy Inference System</i>
ANN	<i>Artificial Neural Networks</i>
ARFIMA	<i>Autoregressive Fractionally Integrated Moving Average</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BM	<i>Brick and Mortar</i>
BPNN	<i>Back-propagation Neural Networks</i>
C2C	<i>Consumer-to-Consumer</i>
CSV	<i>Comma-separated values</i>
DISC	<i>Discounted Combination Forecast Model</i>
DL	<i>Deep Learning</i>
DT	<i>Decision Trees</i>
ELM	<i>Extreme Learning Machines</i>
ERA	<i>Relative Absolute Error</i>
ETS	<i>Exponential Smoothing</i>
GHSOM	<i>Growing Hierarchical Self-organizing Map</i>
HS	<i>Harmony Search</i>
INMET	Instituto Nacional de Meteorologia
IoT	<i>Internet of Things</i>
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>

MASE	<i>Mean Absolute Scaled Error</i>
MIT	<i>Massachusetts Institute of Technology</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MPE	<i>Mean Percentage Error</i>
MS	Microsoft
MSE	<i>Mean Square Error</i>
ND	<i>Normalized Deviation</i>
NRS	<i>Neighborhood Rough Set</i>
OHMA	<i>Ordinary day and Holiday Moving Average</i>
$R^2$	Coefficiente de determinação
$R^2_{\text{ajust.}}$	Coefficiente de determinação ajustado
RMSE	<i>Root Mean Square Error</i>
RNA	Redes Neurais Artificiais
RRSE	<i>Root Relative Squared Error</i>
RVS	Regressão de Vetores de Suporte
SARIMA	<i>Seasonal Autoregressive Integrated Moving Average</i>
sMAPE	<i>Symmetric Mean Absolute Percentage Error</i>
SMO	<i>Sequential Minimal Optimization</i>
SOM	<i>Self-organizing Map</i>
SVM	<i>Support Vector Machines</i>
SVR	<i>Support Vector Regression</i>
U-Theil	<i>Theil's U-statistic</i>
WNN	<i>Wavelet Neural Networks</i>

WT-NNPSO      *Wavelet-Neural Networks-Particle Swarm Optimization-  
Simulation-Optimization*

# LISTA DE TABELAS

<b>Tabela 1.</b> Frequência de medidas adotadas para precisão de previsão de demanda.....	34
<b>Tabela 2.</b> Autores com as medidas de precisão de previsão de demanda adotadas.....	35
<b>Tabela 3.</b> Tabela de correlação entre variáveis.....	64
<b>Tabela 4.</b> Parâmetros do modelo de Rede Neural Artificial.....	72
<b>Tabela 5.</b> MAPE obtido no modelo RNA proposto.....	73
<b>Tabela 6.</b> MAE obtido no modelo RNA proposto.....	75
<b>Tabela 7.</b> RMSE obtido no modelo RNA proposto.....	76
<b>Tabela 8.</b> Resumo comparativo das medidas de performance para previsão de demanda. .....	77

# LISTA DE QUADROS

<b>Quadro 1.</b> Principais funções de ativação. ....	29
<b>Quadro 2.</b> Resumo dos artigos presentes na revisão literária no período de 2010 a 2020. .....	50
<b>Quadro 3.</b> Estatística descritiva dos dados. ....	67

# LISTA DE FIGURAS

<b>Figura 1.</b> Representação da jornada de compras com omnicanalidade. ....	27
<b>Figura 2.</b> Modelo de um perceptron concebido por Rosenblatt. ....	28
<b>Figura 3.</b> Analogia entre um perceptron e um neurônio biológico.....	30
<b>Figura 4.</b> Representação das camadas de uma rede neural artificial. ....	32
<b>Figura 5.</b> Metodologia de pesquisa para proposição e avaliação do modelo de previsão de demanda. ....	61
<b>Figura 6.</b> Evolução de vendas de 2019 a 2020 nos canais online e offline. ....	68
<b>Figura 7.</b> Agrupamento dos produtos em quatro clusters.....	69
<b>Figura 8.</b> Representação dos quatro clusters nos quadrantes de venda online, offline, alcoólicos e não alcoólicos. ....	70
<b>Figura 9.</b> Evolução do MAPE com o aumento do número de neurônios por camada. .	71
<b>Figura 10.</b> Evolução do MAPE por época implementada pelo algoritmo de rede neural artificial.....	72
<b>Figura 11.</b> Evolução semanal de vendas da bebida alcoólica com maior venda.....	78
<b>Figura 12.</b> Evolução semanal de vendas da bebida não alcoólica com maior venda. ...	79

# 1 INTRODUÇÃO

## 1.1 CONTEXTO

Com o desenvolvimento e evolução do mercado de varejo, ligado principalmente à evolução tecnológica, torna possível novos modelos de negócio com integração entre os canais *offline* e *online*. A utilização dos dados disponíveis de mercado e consumidores de diversas fontes, através de modelos de inteligência artificial, assim como definição de estratégias sinérgicas de todos os canais e pontos de contato com os consumidores, consolidam o varejo em um ambiente de omnicanalidade ou *omnichannel*, como apontam Piotrowicz e Cuthbertson (2014); Qureshi *et al.* (2017); Peter *et al.* (2015); Paul *et al.* (2019).

O modelo de um único canal ocorre quando há somente o ponto de venda físico, feito de tijolos e cimento (*Brick-and-Mortar – BM*) ou somente a operação *online* como canal de venda, no qual toda a experiência de compra do consumidor é realizada em um único ponto de interação. Com a adoção da internet, o varejo que inicialmente operava *offline*, passou a operar *online*, podendo ser em uma mesma organização, constituindo multicanais de atendimento, em que o consumidor tem a opção de escolha entre o canal *online* ou *offline*, porém sem integração entre as operações e sem a opção de mudança de canal para uma transação já iniciada pelos consumidores. No início da década de 2010, surge o conceito de omnicanalidade que traz sinergia tanto para as operações, logística e *marketing*, quanto a possibilidade para o consumidor de iniciar a compra em um canal e finalizar em outro (*online/offline*), com diversos pontos de interação para a escolha mais conveniente no momento e local em que estiver (*websites*, redes sociais, loja física, opções de entrega ou coleta, etc.), retirando assim, a barreira do *online* e *offline*,

(ABREU, 2018; BRYNJOLFSSON, 2013; HUBNER, 2016; MATIAS, 2018; RIGBY, 2011).

A complexidade das decisões gerenciais aumenta consideravelmente quando se tem mais de um ponto de interação com o consumidor. Conforme Wang *et al.* (2018); Salinas *et al.* (2019), o produto pode estar fisicamente nas prateleiras ou em um centro de distribuição para atender a demanda dos diferentes canais de forma integrada, exigindo adoção de estratégias precisas de alocação do produto em múltiplos locais.

Marchet *et al.* (2018) argumenta que a omnicanalidade continua a evoluir por uma adaptação pelas empresas considerando: a maturidade da operação, características do produto, necessidade dos consumidores e serviços de conveniência, (KIM, 2018). É relevante, neste meio de disrupção, saber como capturar valor para a cadeia, atender as necessidades dos consumidores e reduzir o tempo do ciclo de compras, ao mesmo tempo que também é necessário manter a lucratividade do negócio.

Para Martino *et al.* (2019), a integração do *online* com o *offline* no varejo, que requer uma resposta ágil nas movimentações na cadeia de suprimentos para atender as expectativas do consumidor, se torna um fator de competitividade. Coles (2017) sugere que a utilização dos dados do comércio eletrônico através de algoritmos de *machine learning* (ML) leva a melhores soluções para os abastecimentos iniciais e ressuprimentos para as lojas físicas em comparação aos modelos tradicionais de previsão de demanda. Matias B. (2018) sugere que a adoção de inteligência artificial na gestão da cadeia de suprimentos é um caminho para diferenciação no mercado e é um valor reconhecido pelos consumidores, tornando-se relevante para melhorar a participação no mercado, conforme sugere Blanco (2018).

Além disso, Qu *et al.* (2017) aponta que o mercado *online* apresenta uma velocidade muito maior com relação às políticas de precificação e promoções, o que deixa em desvantagem as lojas físicas por não terem esta agilidade. O consumidor final ganha mais informações para decisão de compra, tendo um impacto na demanda por canal, forçando os varejistas a estarem sempre um passo à frente na capacidade de analisar e interpretar os movimentos do mercado de forma a antecipar os anseios futuros dos consumidores, como explica Aras *et al.* (2017).

A previsão de demanda se torna então um fator decisivo nas decisões estratégicas, táticas e operacionais da organização, assim defendido por Fildes *et al.* (2019), como a política de capacidade e ressurgimento em que, de forma crescente, mais fatores a influência de forma dinâmica. Um erro em sua estimativa pode gerar um prejuízo e impactar a eficiência operacional da organização, o excesso de custo com inventário, limitação de espaço disponível em prateleiras e a lealdade do consumidor, (CHAWLA; GIRI, 2019). da Veiga *et al.* (2016); Kharfan (2018) apontam que sem a previsão de demanda, a organização consegue agir somente de forma responsiva, gerando perdas em vendas, comprometendo o nível de serviço e má utilização dos recursos (mão de obra, excesso ou falta de inventário e logística). Acrescentam também, Böse *et al.* (2017); Loureiro *et al.* (2018) que a previsão de demanda dos produtos é um dos principais desafios de qualquer varejo. Além de diminuir a energia utilizada na cadeia de suprimentos, com a redução de viagens de transporte, emissão de carbono e outros impactos negativos para o meio ambiente, conforme explorador por Lau (2012).

Dispositivos conectados à internet, ou internet das coisas (IoT – *Internet of Things*) estão espalhados ao longo da jornada de compras, gerando um volume de dados expressivo, que precisam de processamento para se tornarem em informações relevantes

no processo de decisão, tornando a tratativa e análise dos dados uma etapa essencial para o planejamento estratégico.

Em um ambiente de varejo com múltiplos canais disponíveis de forma integrada, a complexidade dos dados para previsão de demanda apresenta características não lineares e sazonalidade nos dados, exigindo modelos como Redes Neurais Artificiais (RNA) e Regressão de Vetores de Suporte (RVS), como citado por Cao e Wu (2016); Semenov *et al.* (2017). Existem ainda, conforme Böse *et al.* (2017), tipos de produtos com comportamento de mercado diferentes, tais como itens com ciclos de vida curto, itens com alta sazonalidade, itens muito sensíveis à promoções e itens com demanda dispersa, o que dificulta a implementação de uma solução única que consiga tratar todos estes grupos, (DA VEIGA, 2016). Além disso, Loureiro *et al.* (2018); Fildes *et al.* (2019) apontam que modelos estatísticos de series temporais, com dados indexados por ordem cronológica, apesar de fácil interpretação e implementação, como ARIMA (*autoregressive integrated moving average*), SARIMA (*seasonal ARIMA*), *exponential smoothing*, *regression*, Box & Jenkins e Holt-Winters possuem limitações como a possibilidade de considerar dados de atributos qualitativos em quantitativos e agregação de dados. Faloutsos *et al.* (2018) reforça ainda, que a transição de modelos tradicionais de previsão de demanda para modelos assistidos por computadores, orientados em dados (*data-driven*), com um certo grau de autonomia e taxa de aprendizado, pode ser atribuída pela disponibilidade de grandes volumes de dados, ricos em múltiplos atributos e com diferentes origens de um mesmo objeto de estudo. Informações de imagens de produtos além de atributos como cor, estilo, *design* e outros, também podem ser significativas para compor a previsão de demanda com adoção de ML, como aponta Giri *et al.* (2019).

O desenvolvimento de soluções com algoritmos integrados de ML ainda estão nos estágios iniciais, sugere Matias B. (2018), e é uma possibilidade promissora de estudos

futuros, conforme Kharfan (2018). Para alcançar os níveis de serviço de forma competitiva, a automação e a inteligência artificial se tornarão mandatárias, aponta Blanco (2018).

A literatura demonstra o interesse em obter modelos com maior previsibilidade de demanda no varejo, utilizando inteligência artificial e dados de fontes diversas e indica que ainda há uma lacuna para novas pesquisas, (MATIAS, 2018; MELACINI, 2018; KHARFAN, 2018). O tema desta pesquisa é o varejo em um ambiente de omnicanalidade, de forma a utilizar informações que se mostrem relevantes para melhorar o nível de serviço para o consumidor através da previsão de demanda por canal, utilizando algoritmo de ML com múltiplas variáveis ao longo da cadeia do varejo e auxiliar na tomada de decisões gerenciais, tornando assim um fator que agregue mais valor para o consumidor final se compararmos com métodos de gestão tradicionais.

## 1.2 PROBLEMÁTICA

O varejo que oferece ao cliente somente um canal de vendas (*online* ou *offline*) tem uma política de gestão de suprimento muito mais simples do que em comparação quando há integração de omnicanalidade, em que o cliente pode decidir, comprar, receber, pagar, revisar e devolver o produto em canais distintos. Com esta integração, a complexidade da cadeia de suprimentos, bem como as políticas e decisões sobre onde, quando e quanto estocar os produtos passam a exigir cada vez mais das operações de varejo. Por outro lado, as informações sobre os consumidores estão cada dia mais abundantes e saber utilizar estes dados para a tomada de decisão passa a ser um fator de competitividade no mercado. Portanto, saber lidar com grandes volumes de dados de forma rápida e precisa é um grande desafio para o ambiente acadêmico e corporativo.

Este cenário de omnicanalidade, por se apresentar muito dinâmico e influenciado por fatores internos (demanda histórica, preço, margem, custos operacionais, etc.) e externos (perfil do cliente, localização do ponto de interação, clima, concorrentes, marketing, etc.), indica que um modelo com múltiplas variáveis seja adequado para representá-lo. Ao mesmo tempo que a análise de um grande volume de dados pode se mostrar um passo importante no processo de decisão na gestão de suprimentos através de uma previsão de demanda que utilize múltiplos fatores na sua composição.

Os modelos econométricos apresentam a necessidade de tratativa dos dados para verificar as suposições de linearidade, colinearidade, homoscedasticidade, normalidade e independência através de transformações e retirada de valores discrepantes. Isto é dificultado em dados complexos e não lineares, o que indica, para estes cenários de informações, a adoção de um método dirigido por dados, em que estas relações são estipuladas pelos algoritmos de *machine learning*, sem a necessidade de tratativa nos dados de forma tão rigorosa quanto exigida em modelos econométricos.

### 1.2.1 Questão geral

Este trabalho pretende responder a seguinte questão, com suporte de uma revisão da literatura:

- Como os dados internos (histórico de vendas, data, categoria mercadológica, característica do produto, canal de vendas, precificação interna) e dados externos (dia da semana, feriados, precipitação, temperatura do ambiente) podem ser utilizadas para auxiliar na previsão de demanda em um cenário de varejo com omnicanalidade?

### 1.2.2 Questões específicas

Cada ponto de interação com a jornada de compras deverá ter fatores de maior relevância para tomada de decisão sobre quais produtos deverão ser estocados ou expostos nas lojas, bem como a necessidade de ressuprimento, suportados pela previsão de demanda por canal. O modelo proposto deverá ser *data-driven* e adaptativo a cada cenário com diferentes atributos a serem analisados.

Temos portanto, as seguintes questões de pesquisa que irão guiar este trabalho.

- Quais são os algoritmos mais frequentes de inteligência artificial abordados pelos modelos presentes na literatura para previsão de demanda para o varejo com omnicanalidade?

- Quais são as métricas de erros para avaliação e comparação entre os modelos propostos?

- Como realizar previsão de demanda no varejo com omnicanalidade utilizando os dados internos e externos através de um modelo de inteligência artificial?

### 1.3 OBJETIVOS DA PESQUISA

Os objetivos apresentados a seguir foram formulados a fim de responder as questões de pesquisa.

### 1.3.1 Objetivo geral

Identificar os modelos presentes na literatura dos últimos dez anos em que são discutidos o tema de omnicanalidade no varejo, que envolvem previsão de demanda com adoção de inteligência artificial em um cenário de varejo em caminho à omnicanalidade. Este período foi escolhido pelo termo de omnicanalidade ter sido concebido e pesquisado pela literatura desde o início da década de 2010 (RIGBY, 2011).

Posteriormente, desenvolver um método de análise de dados para elaboração de um modelo adaptativo e preditivo para o varejo com omnicanalidade, que permite considerar variáveis dinâmicas que dê um melhor suporte para definição de uma política de ressuprimento dos produtos ao longo da cadeia de suprimentos, com base na previsão de demanda gerada pelo modelo proposto em comparação a um modelo tradicional.

### 1.3.2 Objetivos específicos

Os objetivos deste trabalho são:

- i) Utilizar atributos internos e externos para a previsão de demanda;
- ii) Realizar a segmentação dos dados conforme a sua similaridade através de um algoritmo de clusterização (*Hierarchical Clustering*);
- iii) Executar uma Rede Neural Artificial para previsão de demanda nos dados clusterizados.

## 1.4 JUSTIFICATIVA

Os atuais algoritmos de ML estão em uma maturidade de resolução de problemas bem complexos, já presentes em nosso cotidiano em publicidades personalizadas, filtros de e-mail, análise de risco, identificação de humor, reconhecimento de voz e face, diagnóstico de doenças, categorização de perfis similares, sugestão de filmes, etc. Porém, a adoção no varejo com omnicanalidade ainda está em seus primórdios, conforme Matias B. (2018). A construção de um modelo que considere o dados disponíveis e relevantes para treinar estes algoritmos se torna uma oportunidade de pesquisa relevante para o meio acadêmico e com forte interesse em aplicações nas corporações.

O uso dos dados disponíveis e relevantes no gerenciamento da cadeia de suprimentos passa a ser um fator de competitividade e não mais um diferencial. A necessidade de tratar grandes volumes de dados para elevar o nível de serviço do varejo com omnicanalidade se torna cada vez mais relevante. A capacidade de identificação dos fatores que mais influenciam a previsão de demanda presentes nos dados através de algoritmos de inteligência artificial se apresenta um conhecimento de importância tanto para a academia quanto para o mercado.

A previsão de demanda é um fluxo de informação que influencia os fluxos financeiros e de movimentação física em toda cadeia de suprimentos para no planejamento desde a quantidade de matéria prima, lotes de produção, definição de mão de obra, logística de envio, até o atendimento ao consumidor final. Portanto, a busca por um modelo de previsão de demanda mais precisa que os modelos tradicionais, é justificada tanto pela redução de custo em toda a cadeia de suprimento quanto pela melhora no nível de serviço percebido pelo consumidor, em que a disponibilidade do

produto desejado no canal certo e momento certo reduzem o tempo total da compra e reduz desistências de consumo.

## 1.5 SÍNTESE METODOLÓGICA

Para alcançar os objetivos do estudo, a investigação se iniciou com uma identificação e revisão dos artigos científicos que pesquisam a omnicanalidade, que se iniciou em 2010, de forma a garantir uma sustentação teórica à pesquisa. Em seguida, foi estabelecida uma parceria com um estabelecimento comercial com presença de omnicanalidade no setor de varejo alimentar na cidade de Goiânia / GO, para a obtenção do acesso a um banco de dados com registros de transações de vendas de um período de 2 anos. A terceira etapa foi a escolha do método de análise dos dados, através da construção do modelo de algoritmos de inteligência artificial (*hierarchical clustering* e redes neurais artificiais). Por fim, é apresentada a análise dos dados e comparativo entre o modelo adotado e a literatura.

## 1.6 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em quatro capítulos após a introdução. No capítulo 2, é realizada uma revisão da literatura para identificar: (i) as principais metodologias que abordam a previsão de demanda no varejo em omnicanalidade com adoção de inteligência artificial, (ii) lacuna da literatura, (iii) definição de conceitos e terminologia de referência, (iv) identificação das principais métricas de avaliação do modelo preditivo. No capítulo 3 é apresentada a abordagem metodológica, com a delimitação da pesquisa, etapas da

realização da investigação, ferramentas utilizadas e resultados esperados. Em sequência, no capítulo 4 são apresentadas as variáveis coletadas e implementação do modelo proposto com a comparação através de métricas de desempenho com a literatura. Por fim, o capítulo 5 trata sobre as conclusões, considerações finais e dificuldades encontradas na implementação dos modelos e sugestões para trabalhos futuros.

## 2 REVISÃO DA LITERATURA

Conforme Berman (2019) a jornada de compras com omnicanalidade passa pela descoberta inicial do produto, busca por informações, compra, pagamento, entrega e possível devolução do produto, oferecendo aos consumidores múltiplos canais de interação antes, durante e depois da compra, proporcionando uma experiência de compra única (MATIAS, 2018). Estes canais podem ser *websites*, redes sociais, aplicativos móveis, quiosques, correspondências, catálogos, televendas, consoles de jogos, televisão e muitos outros, como aponta Rigby (2011), cada um fornecendo informações e garantindo o suprimento nos canais de preferência de cada consumidor (BELL, 2015; FAIRCHILD, 2014).

Conforme a experiência no varejo aproxima os canais de venda com as lojas físicas e virtuais, Erik Brynjolfsson, Yu Jeffrey Hu (2013) há a indicação de que a cadeia de suprimento está em um momento de transformação disruptiva. A omnicanalidade propõe uma experiência de serviços unificados, com baixa variabilidade e alto nível de satisfação através de disponibilidade das informações, possibilidades de interação com os consumidores e diferentes formatos de entrega do produto final, (KUMAR, 2019). Portanto, Erik Brynjolfsson, Yu Jeffrey Hu (2013) sugere que a reformulação das estratégias que envolvem cadeia de suprimentos se torna crítica para que o varejo mantenha sua competitividade. Além disso, a gestão de relacionamento com o consumidor deve adotar ferramentas que incluem comércio eletrônico, redes sociais, interação com comunidades virtuais, transações móveis, análise de dados, algoritmos de inteligência artificial e realidade aumentada, tendendo a menos interações entre humanos, observando maiores distâncias e sem restrição de tempo (STEINHOFF, 2019).

Progressivamente, mais informações sobre consumidores estão disponíveis, geradas por dispositivos eletrônicos conectados à internet, sendo que a quantidade de

dados dobra a cada dois anos (MATIAS, 2018). Por outro lado, consumidores estão recebendo mais propagandas e sugestões de compras que, em maior parte, se tornam inefetivas. Visto que o volume de dados é enorme, a sua análise não se torna prática em planilhas eletrônicas. Deste modo, algoritmos de ML são mais precisos na identificação de consumidores, customizando estratégias de marketing e performando melhor o retorno do investimento em comparação com modelos tradicionais (THOBANI, 2018).

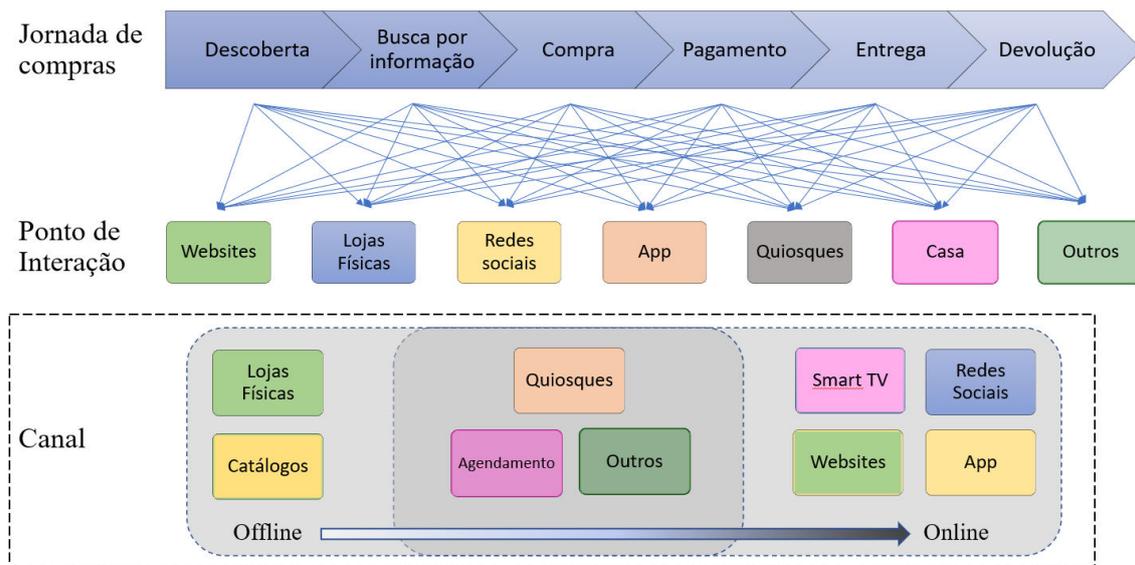
Uma sugestão de compra com sucesso é aquela que reduz o tempo de procura do cliente durante a jornada de compras e que aumenta as vendas. Existem diferentes tipos de recomendação: itens que foram previamente visualizados por outros consumidores, itens que estão em tendência de aumento de vendas, itens que foram visualizados por outros consumidores, itens similares e complementares, sistema de recomendação de outros clientes. A previsão através de ML pode considerar múltiplos atributos como preço, nome do produto, reconhecimento de imagem, variação de produtos, clubes de compra, histórico de compras, *rankings*, disponibilidade de estoque, informações de geolocalização, clima, idade, gênero, estado civil e outros (THOBANI, 2018).

## 2.1 OMNICALIDADE

O processo de compras é dado através da jornada de compras que se inicia pela descoberta do produto, busca de informação, compra, pagamento, entrega e uma possível devolução (SAGHIRI, 2017). Esta jornada é composta pelo fluxo de dados, fluxo financeiro e fluxo de mercadorias que podem ocorrer em diferentes canais de forma independente, seja *offline* (lojas físicas, catálogos, etc.) ou *online* (*websites*, aplicativos móveis, redes sociais, etc.) que estão integrados.

Cada um destes canais apresentam diferentes perfis de compras, lucratividade e lealdade de consumidor (PETER, 2015). Nos canais pode haver interações simultâneas, em que o consumidor tem a possibilidade de conhecer o produto na loja física e comparar os preços e referências de outros consumidores *online* antes de tomar a decisão de compra. Ao contrário, o consumidor também poderá conhecer as informações do produto como marca, preço e revisões online pela internet e ir à loja física somente para ter a experiência de conhecer a qualidade o produto fisicamente. Nesta troca de canais ao longo da jornada de compras, de forma integrada, harmoniosa e complementar, temos um cenário de omnicanalidade com a capacidade de atender a demanda em qualquer hora e local que o consumidor estiver (BELL, 2015; BERMAN, 2019; FAIRCHILD, 2014; MATIAS, 2018; RIGBY, 2011). Em um contexto de omnicanalidade, há integração da operação, logística, inventário e troca de informações para convergência dos processos de ressuprimento (HÜBNER, 2016).

Esta interatividade entre o *offline* e *online* é percebida por toda a jornada de compras conforme ilustrado pela figura 1. Nela o consumidor pode iniciar a jornada com a descoberta *offline*, transitar *online* para obter mais informações, realizar a compra no canal e ponto de interação de preferência, efetuar o pagamento através de aplicativos ou fisicamente na loja, escolher a retirada do produto em algum ponto de preferência ou solicitar entrega e realizar a devolução da forma mais conveniente. Existem, portanto, diversos caminhos possíveis que estão disponíveis para escolha em um ambiente de omnicanalidade, em que a experiência e o nível de serviço ao longo da cadeia é o que agrega valor ao consumidor.



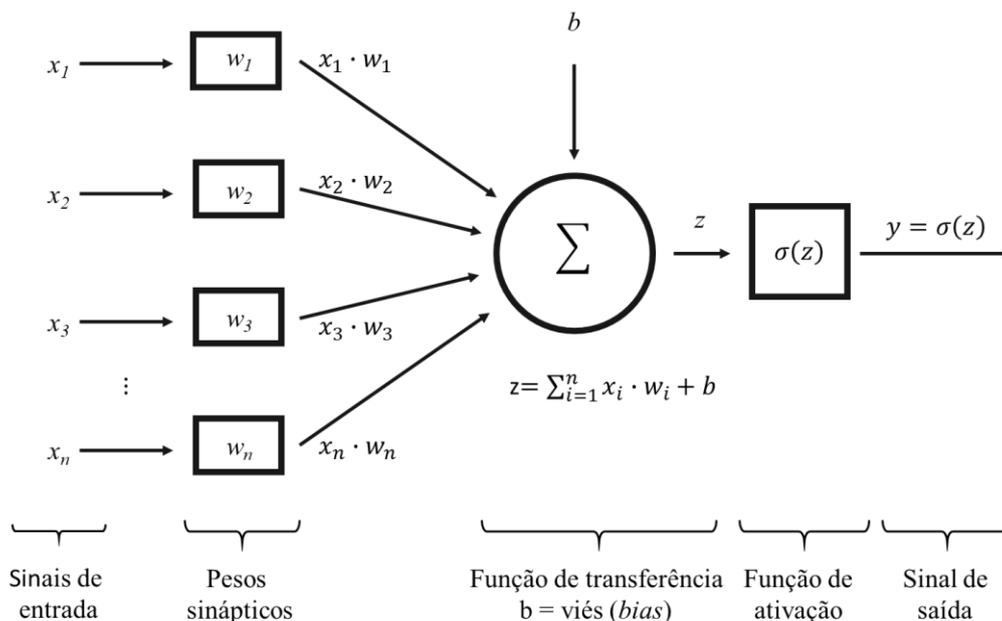
**Figura 1.** Representação da jornada de compras com omnicanalidade.

**Fonte:** Elaborado pelo autor (2021)

## 2.2 MACHINE LEARNING E REDES NEURAIS ARTIFICIAIS

Dutton e Conroy (1996) define *Machine Learning* (ML) como o processo de melhorar uma habilidade de realizar uma tarefa através de computadores, seja através de dados externos ou tornando mais efetivo o conhecimento já adquirido, sem que seja realizado através de uma programação direta.

O ML contém um algoritmo que representa um neurônio biológico e sua relação contígua com demais neurônios que formam uma rede neural. Este é o modelo concebido por Rosenblatt (1957), denominado de *perceptron*, que recebe informações externas (dados), os processa e devolve um dado de saída. O *perceptron*, representado pela figura 2, é um modelo supervisionado e binário, ou seja, para os dados processados, sabemos o resultado esperado para confrontar com o valor binário previsto (0 ou 1).



**Figura 2.** Modelo de um *perceptron* concebido por Rosenblatt.

**Fonte:** Elaborado pelo autor (2021)

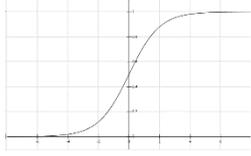
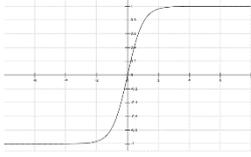
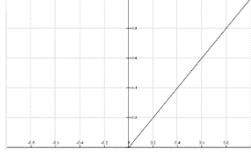
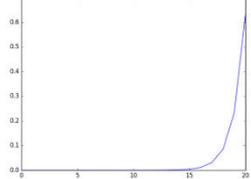
Em um *perceptron*, os sinais de entrada  $x_1, x_2, x_3, \dots, x_n$ , correspondem aos atributos ou variáveis a serem utilizadas pelo modelo, em que cada um terá um peso ( $w_1, w_2, w_3, \dots, w_n$ ) que serão somados, juntamente com o viés  $b$  através da função de transferência a seguir descrita.

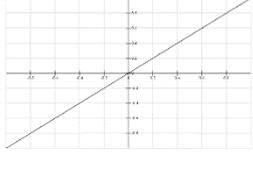
$$z = \sum_{i=1}^n x_i \cdot w_i + b \quad (4)$$

Em seguida é processado pela função de ativação  $\sigma(z)$ , gerando o sinal de saída do modelo preditor,  $y$ , Bielecki (2019). O termo  $b$ , ou viés, age como o coeficiente linear, possuindo o efeito de aumentar ou diminuir o valor da função de ativação (HAYKIN, 2009) e corresponde ao valor de saída do neurônio, caso a entrada seja zero. Com o viés, a rede é capaz de resolver problemas de maior complexidade em relação aos casos em que se ele fosse ausente.

Até este ponto, o neurônio aplicou uma transformação linear através da função de transferência. O próximo passo é aplicar a função de ativação que, geralmente, é uma função que agrega um grau de não linearidade e age como o ativador do sinal de saída, tornando a rede neural capaz de aprender padrões complexos presentes nos dados. Uma rede neural sem a função de ativação se torna um modelo de regressão linear. As principais funções de ativação estão apresentadas no quadro 1 (HANSEN, 2019).

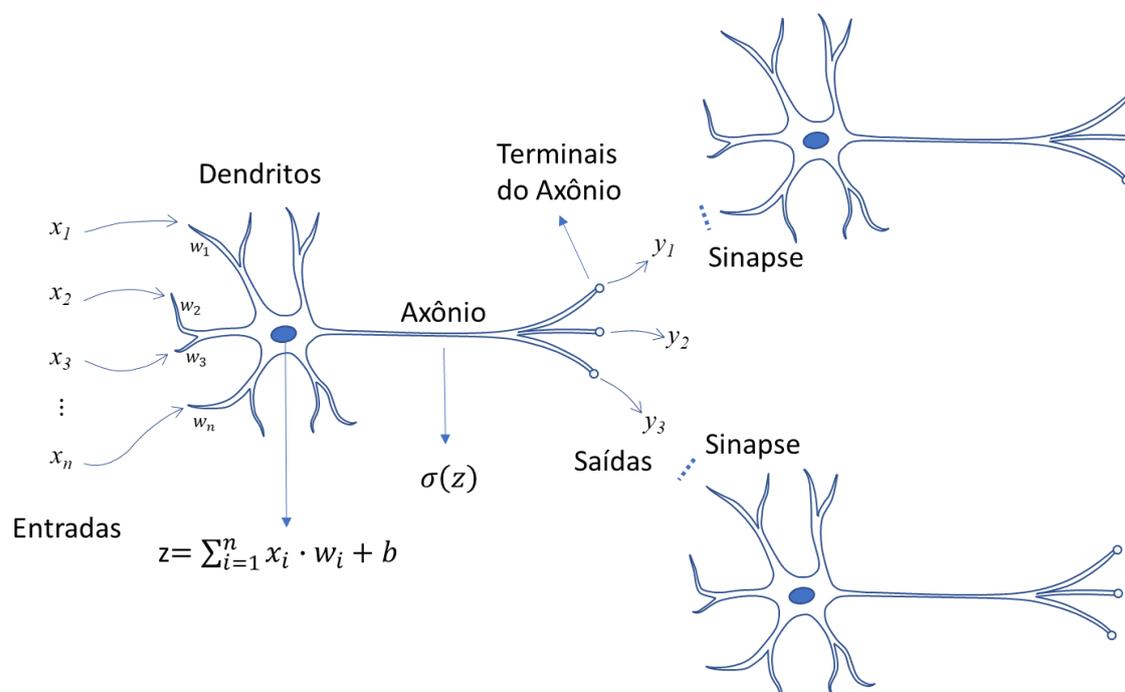
**Quadro 1.** Principais funções de ativação.

Função de Ativação	Equação	Gráfico	Observação
Sigmoide ou Logística	$\sigma = \frac{1}{1 + e^{-x}}$		Manipulação simples e rápida, porém com perda de informação a cada camada processada, devido sua derivada ter curto alcance. É uma função utilizada para classificação binária.
Tangente Hiperbólica	$\tanh = \frac{2}{1 + e^{-2x}} - 1$		Semelhante à sigmoide, porém com a saída para valores negativos e também com saída próximo a 0 para entradas igual a 0. Geralmente utilizada para classificação de duas classes.
ReLU	$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } x \geq 0 \end{cases}$		Utilizado em <i>Deep Learning</i> para visão computacional e reconhecimento de fala. Sua implementação é mais rápida que a função sigmoide. Possui o problema de morte de alguns neurônios, independente dos sinais de entrada.
Softmax	$\sigma(y)_i = \frac{e^{y_i}}{\sum_{j=1}^J e^{y_j}}$ para $i = 1, 2, 3, \dots, J$		A somatória de suas saídas é sempre 1, configurando uma distribuição de probabilidade. Geralmente utilizada na última camada da rede neural como função de classificação múltipla.

Identidade	$f(x) = x$		Não trata a complexidade dos dados que geralmente alimentam a rede neural.
------------	------------	---	--

No modelo RNA proposto neste estudo, foi adotado a função de ativação sigmoide, o que transforma as somas sinápticas lineares, em cada neurônio, em uma função não linear (SEMENOV, 2017; WANG, 2018), o que permite que o algoritmo aprenda relações lineares e não lineares entre os dados de entrada e saída (SALAMI, 2018; VERSTRAETE, 2019).

Em analogia a um neurônio biológico, o *perceptron* é um modelo simples para sua representação, em que os pesos  $w_n$  correspondem aos dendritos, a função de ativação ao axônio, os sinais de saída correspondem aos terminais do axônio e os arcos de ligação com os próximos neurônios da rede correspondem às sinapses, figura 3.



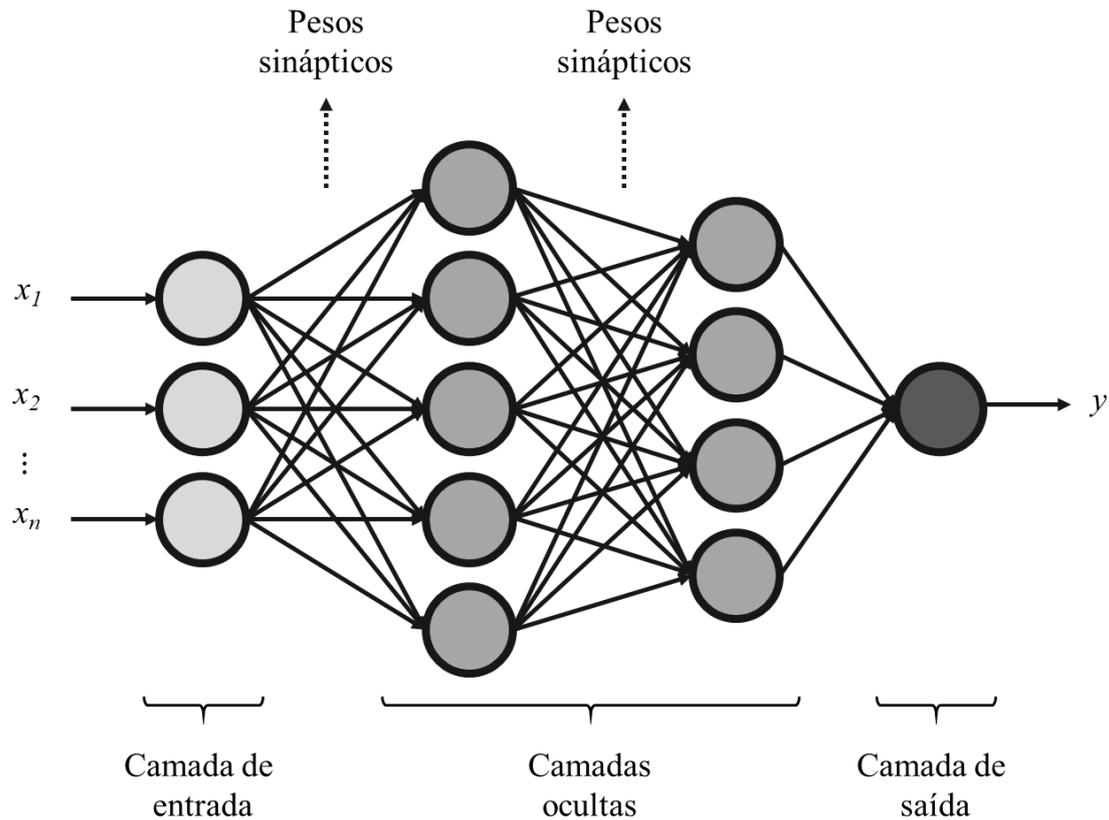
**Figura 3.** Analogia entre um *perceptron* e um neurônio biológico.

**Fonte:** Elaborado pelo autor (2021)

O objetivo do algoritmo *perceptron* é treinar com o conjunto de dados de forma supervisionada, ou seja, com a classe dos sinais de saída conhecidos, de modo a estabelecer os pesos sinápticos, de maneira interativa, para atender a previsão de saída  $y$  para novos dados não pertencidos aos dados utilizados para o treinamento.

Em modelos de previsão são utilizados mais de um *perceptron*, de forma a garantir uma melhor adaptação do modelo aos problemas reais com presença de ruído, formando assim uma rede neural em que é descrita por nós de processamentos (neurônios), interconectados de acordo com o modelo desejado, realizando cálculos numéricos localmente através da função de ativação e formando sinais de saída transformados para outros neurônios da rede (ADAMOWSKI, 2012). As conexões entre os neurônios através dos pesos sinápticos é a forma de reter o conhecimento adquirido através das interações da rede com os dados (HAYKIN, 2009).

Um *perceptron* com uma única camada possui uma limitação de classificação apenas em dados linearmente separáveis. Para superar esta limitação, o modelo de rede neural com mais camadas de neurônios é descrito pelo modelo Perceptron de Múltiplas Camadas (*Multilayer Perceptron* - MLP), em que os neurônios estão organizados em camadas de forma que os neurônios estão conectados somente com os neurônios da camada imediatamente anterior, contendo pelo menos uma camada oculta. As camadas ocultas têm este nome pelo fato de serem as partes da rede neural que não estão visíveis olhando diretamente para a camada de entrada e camada de saída. Este modelo de fluxo unilateral das informações no sentido da camada de entrada, camadas ocultas, camada de saída é conhecido como *Forward Propagation*. O modelo MLP está ilustrado pela figura 4.



**Figura 4.** Representação das camadas de uma rede neural artificial.

**Fonte:** Elaborado pelo autor (2021)

Como no processo supervisionado é conhecido o valor real da saída desejada, é possível calcular o erro da saída gerada através da função de perda.

$$\varepsilon = (y_{real} - y_{previsto})^2 \quad (5)$$

Em seguida, deve-se atualizar os pesos e vieses da rede, propagando camada a camada no sentido inverso, através de um processo conhecido como retro propagação ou *back propagation*, conforme demonstrado por Rumelhart, Feldman e Hayes (1986), de forma a minimizar o erro resultante de cada neurônio. Para isto, é necessário calcular o gradiente dos parâmetros de cada neurônio individualmente e corrigir os parâmetros conforme a equação.

$$p_{i,j} = p_{i,j} - \eta * \delta p_{i,j} \quad (6)$$

Em que  $p_{i,j}$  é o parâmetro  $i$  da camada  $j$ ,  $\eta$  é a taxa de aprendizado e  $\delta p_{i,j}$  é o gradiente do parâmetro calculado individualmente. A taxa de aprendizado é a proporção que os gradientes são utilizados para atualizar os parâmetros  $p_{i,j}$ .

Durante a atualização dos parâmetros da rede neural (pesos sinápticos, vieses e taxa de aprendizado), é comum a utilização de algoritmos de otimização para minimizar a função de perda  $\varepsilon$ . O algoritmo ADAM (*adaptive moment estimation*) é um método de otimização estocástica descrito por Kingma e Ba (2015) que se mostrou eficaz em modelos de ML como regressão logística, redes neurais e *Deep Learning* para aplicação de problemas reais.

A interação completa de *forward propagation* e *backpropagation* dos dados é conhecida como uma época ou *epoch*, em que durante o processo de aprendizado esta interação ocorre diversas vezes até que se estabilize a taxa de aprendizado.

Conforme Zhang *et al.* (1998), as dificuldades encontradas para a construção de um modelo de redes neurais é a definição da quantidade de neurônios em cada camada, a quantidade de camadas ocultas, quais funções de ativação utilizar, o algoritmo de treinamento, o método de normalização dos dados, a definição da segmentação dos dados para treinamento e teste e a definição dos medidores de performance. Para a definição destes parâmetros, de forma a garantir uma arquitetura de rede neural otimizada, não se tem um método único, seja de forma heurística ou através de simulação.

O objetivo de uma rede neural é a generalização do modelo de forma a obter resultados de saída aceitáveis para dados de entrada que não foram utilizados durante o treinamento, mas que fazem parte da mesma população da qual os dados para o treinamento foram extraídos (HAYKIN, 2009). Uma rede bem treinada é aquela que pode ser colocada na prática de um problema real (GOH, 1995).

## 2.3 MÉTRICAS DE DESEMPENHO

A metodologia compara a precisão de previsão de demanda geradas através do modelo de Redes Neurais Artificiais (RNA) com o resultado observado na literatura. As métricas de desempenho foram escolhidas conforme 78,1% de utilização presente na revisão bibliográfica, conforme a apresentado na tabela 1, sendo elas MAPE, RMSE, MSE e MAE.

**Tabela 1.** Frequência de medidas adotadas para precisão de previsão de demanda.

Métrica de Desempenho	Frequência de Utilização	% Frequência	% Freq. Acum.
<b>MAPE</b> ( <i>Mean Absolute Percentage Error</i> )	<b>17</b>	<b>27,0%</b>	<b>27,0%</b>
<b>RMSE</b> ( <i>Root Mean Square Error</i> )	<b>14</b>	<b>22,2%</b>	<b>49,2%</b>
<b>MSE</b> ( <i>Mean Square Error</i> )	<b>11</b>	<b>17,5%</b>	<b>66,7%</b>
<b>MAE</b> ( <i>Mean Absolute Error</i> )	<b>7</b>	<b>11,1%</b>	<b>77,8%</b>
MASE ( <i>Mean Absolute Scaled Error</i> )	3	4,8%	82,5%
R <sup>2</sup> ( <i>Coefficient of Correlation</i> )	3	4,8%	87,3%
U-Theil ( <i>Theil's U-statistic</i> )	2	3,2%	90,5%
sMAPE ( <i>Symmetric MAPE</i> )	2	3,2%	93,7%
MPE ( <i>Mean Percentage Error</i> )	1	1,6%	95,2%
RAE ( <i>Relative Absolute Error</i> )	1	1,6%	96,8%
ND ( <i>Normalized Deviation</i> )	1	1,6%	98,4%
RRSE ( <i>Root Relative Squared Error</i> )	1	1,6%	100,0%
<b>Total Geral</b>	<b>63</b>	<b>100,0%</b>	

Os autores com suas respectivas medidas adotadas estão apresentados na tabela dois.



2020	Caglayan N., Satoglu S.I., Kapukaya E.N.	X	X	X										3
<b>Total</b>		<b>18</b>	<b>14</b>	<b>11</b>	<b>7</b>	<b>3</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>64</b>
<b>Geral</b>														

É possível observar que a adoção de apenas uma métrica é realizada por treze dos trinta autores, enquanto nove autores utilizam duas métricas, quatro autores utilizam três métricas e somente quatro autores utilizam quatro ou mais métricas em seus artigos.

### 2.3.1 MSE

A média do erro quadrático (*Mean Square Error* - MSE) é uma medida utilizada para avaliar a performance de modelos de regressão linear (GIRI, 2019). Porém como o erro está elevado ao quadrado, ele se torna muito sensível à medida em que o erro aumenta. Sua definição é dada pela Equação 7.

$$MSE = \frac{1}{n} \sum_{i=1}^n e_n^2 \quad (7)$$

Na qual  $e_n$  é o erro da diferença entre o valor previsto ( $Y_{n,previsto}$ ) e o valor real ( $Y_{n,real}$ ).

### 2.3.2 RMSE

A raiz quadrada da média do erro quadrático (*Root Mean Square Error* - RMSE) é uma medida de fácil aplicação e largamente utilizada em modelagens estatísticas (WONG, 2013). Esta medida pode descrita pela Equação 8.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_n^2} \quad (8)$$

### 2.3.3 MAPE

A média absoluta do erro percentual (*Mean Absolute Percentage Error* - MAPE) é um critério menos sensível a erros maiores como o RMSE (WONG, 2013). Sua expressão pode ser descrita pela Equação 9.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{e_n}{Y_{n,real}} \right| \quad (9)$$

Na qual  $Y_{n,real}$  representa o valor conhecido da n-ésima saída observada.

### 2.3.4 MAE

A média absoluta do erro (MAE) explica a proximidade da previsão com os dados reais com interpretação simples, quanto mais próximo de zero, melhor (GIRI, 2019). Sua representação é definida pela Equação 10.

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_n| \quad (10)$$

Ferreira *et al.* (2016) defende que a variação percentual do MSE pode ser utilizada como um indicador de relevância da variável ou atributo analisado com relação a previsão de demanda proposta pelo modelo.

O MAE e o MSE são apontados por Baquero e Mosquera-Recalde (2020) como os melhores indicadores para comparação de modelos de otimização com presença de demanda intermitente, porém para previsão de demanda com modelos de redes neurais, a utilização somente destes dois não é apropriada.

Conforme indica Huber *et al.* (2019), tanto o MAE quanto o RMSE são medidas de erro dependentes da escala. Portanto, não podem ser utilizadas para comparar séries temporais com dados em escalas diferentes. O mesmo não ocorre para medidas de erro

percentuais como o MAPE, que pode ser utilizado para comparar bases de dados diferentes e com escalas diferentes, como reforçado por Aras *et al.* (2017).

Ainda de acordo com Aras *et al.* (2017) e Huber *et al.* (2019), não existe um consenso sobre qual seria o melhor medidor de desempenho para um modelo de previsão de demanda. Entretanto, quando se deseja comparar a performance de diferentes modelos, *datasets* e escalas, é recomendado utilizar um medidor adimensional como o MAPE.

## 2.4 METODOLOGIAS PARA PREVISÃO DE DEMANDA NA LITERATURA

De acordo com Böse *et al.* (2017), podemos separar o estudo de previsão de demanda em dois grupos:

- i) Metodologia de séries temporais, e.g., ARIMA, *Box & Jenkins*, *Holt-Winters* e *exponential smoothing*, fundados com estatística e econométrica. Estas metodologias já estão bem exploradas na literatura e possuem os seguintes objetivos: extrapolação (diferença de tempo entre o treinamento e teste do modelo), dinâmica temporal e distribuições preditivas. A maioria destes métodos são pobres em dados, de forma que o aumento no volume de dados em pouco interfere o resultado da previsão e possuem poucos atributos para serem aprendidos.
- ii) Métodos supervisionados de ML como classificação e regressão que aprendem com os dados de treinamento e são avaliados em dados de teste, são ricos em dados e podem incrementar o aprendizado com múltiplas fontes de dados (e.g., preço, datas festivas, marca, cor, clima, etc.). Os algoritmos de aprendizado de adaptam facilmente à dinâmica dos dados.

Os estudos que utilizam algoritmos de inteligência artificial para previsão de demanda, em sua maioria, fazem uma comparação entre algoritmos dos dois grupos acima descritos. Os artigos que utilizaram atributos exclusivamente internos à organização são discutidos na seção 2.4.1, enquanto os que utilizaram atributos internos e externos são discutidos na seção 2.4.2.

#### 2.4.1 ARTIGOS COM ATRIBUTOS EXCLUSIVAMENTE INTERNOS

Wong e Guo (2010), (2013) realizaram uma pesquisa para previsão de demanda considerando o horizonte de médio prazo no mercado de moda. A pesquisa teve como objetivo promover melhorias na cadeia de suprimentos através de um modelo híbrido de redes neurais *Extreme Learning Machines* (ELM), integrado com um processo heurístico, a fim de promover uma habilidade de generalização da rede neural, denominado por *Hybrid intelligent forecaster*. Em um modelo ELM de rede neural *feedforward* os pesos sinápticos das camadas ocultas podem ser gerados de forma aleatória com transformações não lineares, sem a necessidade de serem ajustados. Os parâmetros que o pesquisador deverá determinar são os pesos sinápticos da camada de saída, o que torna o modelo mais rápido no treinamento, em comparação com redes neurais com *backpropagation* Huang *et al.* (2006). Nas pesquisas de Wong *et al.* (2010), (2013) foram realizados três experimentos de quatro categorias de produtos (saia, jaqueta, casaco e calça), com granularidades de previsão mensal, trimestral e anual. As médias do medidor MAPE obtidas com os dois modelos (proposto e ARIMA) para as quatro categorias de produtos foram, respectivamente, 35,0% (previsão mensal), 15,1% (previsão trimestral), 12,38% (previsão anual), 1692,7% (ARIMA, previsão mensal) e 25,8% (ARIMA, previsão

trimestral). No modelo ARIMA os dados analisados foram insuficientes para a previsão anual.

É possível observar que o modelo proposto se torna mais preciso à medida que o nível de agregação dos dados de demanda aumenta do período mensal para anual. Além disso, foi obtido um resultado significativamente melhor do modelo proposto em comparação com o modelo ARIMA.

Também com o foco no mercado de moda, Loureiro *et al.* (2018) adotaram um modelo com abordagem de *Deep Neural Network* (DNN) para previsão individual de novos produtos a serem lançados em futuras temporadas, utilizando dados de compras na operação de uma companhia privada. Em sua pesquisa, Loureiro *et al.* (2018) compararam o MAPE obtido com o modelo DNN com outros modelos de *Machine Learning*, como *Decision Trees* (DT), *Random Forest* (RF), SVR, ANN e Regressão Linear (RL). Os dados disponíveis para aplicação dos modelos contemplam diferentes naturezas como, características físicas dos produtos (família, subfamília, tipo de cor, cor, preço e tamanho) e atributos que representam conhecimento de especialistas do setor (classificação de sazonalidade, segmento, tipo de loja e expectativa de vendas), além dos dados históricos de vendas de cada um dos 684 itens avaliados. A performance obtida durante a fase de teste dos modelos DNN, DT, RF, SVR, ANN e RL resultam nos respectivos valores de MAPE: 37,8%, 37,7%, 34,5%, 39,3%, 38,7% e 45,1%. Em conclusão, a adoção de um modelo de DNN, apesar de se apresentar um modelo viável de aplicação em dados pequenos e de grande escala, possui uma dificuldade adicional devido ao complexo processo de treinamento. Neste estudo o modelo DNN obteve resultado de MAPE inferior ao modelo de simples aplicação, RF. Por fim, o autor sugere a adoção de dados externos à organização como dados de concorrentes e redes sociais em conjunto com técnicas de mineração de texto para melhorar a performance do modelo de

predição de vendas para novos produtos no mercado de moda, além da possibilidade da aplicação do modelo DNN para produtos contínuos que possuam um histórico de vendas conhecido.

Da Veiga *et al.* (2016) também compararam a performance do modelo linear clássico ARIMA com dois modelos não lineares propostos (*Wavelets Neural Networks – WNN* e *Takagi-Sugeno Fuzzy System – TS*) utilizando dados agregados de três categorias de produtos em um varejo de alimentos perecíveis para previsão de demanda. Com a utilização dos dados históricos para treinar e testar os modelos, foi demonstrado os seguintes resultados na medição do MAPE para as categorias A, B e C, respectivamente: 1) Modelo TS: 5,73%, 5,72% e 5,51%; 2) Modelo WNN: 5,46%, 4,88% e 4,63%; 3) Modelo ARIMA: 6,52%, 7,23% e 8,57%. Primeiramente, os autores ressaltam que, apesar dos dados serem de uma multinacional varejista com muitas lojas físicas, para utilização nos modelos, estes dados foram agregados em categorias A, B e C. Com isto, a previsibilidade da demanda sofreu menos variações. Nos dados também foram detectados sinais claros de sazonalidade e tendência de crescimento nas vendas com o tempo. Apesar de atingir uma precisão maior nas previsões do modelo WNN, ele possui uma barreira computacional na adoção por empresas, pois exige um tempo significativamente maior para obter os resultados em comparação aos outros modelos presentes no artigo, além de exigir uma experiência maior dos pesquisadores na definição dos parâmetros iniciais da rede neural.

Para trabalhar também de forma agregada, Islek e Oguducu (2015) propuseram agrupar diferentes centros de distribuição com características similares de vendas, para em seguida adotar um modelo combinando a média móvel com um algoritmo de redes bayesianas em cada um dos *clusters*, de forma a se obter uma melhoria na previsão de vendas por produto e centro de distribuição. Os atributos presentes nos dados referentes

ao centro de distribuição são: localização, capacidade, quantidade de veículos, vendas semanais de produtos, área em m<sup>2</sup> da região de vendas, quantidade de empregados e quantidade de clientes; e aos produtos: categoria, quantidade de vendas, tempo de venda.

Como resultado, o MAPE obtido com um modelo simples de média móvel, com o modelo proposto sem clusterização, com clusterização dos centros de distribuição e, por fim, com clusterização dos sub centros de distribuição, foram de 129%, 49%, 24% e 17%, respectivamente. Os autores ressaltaram que, para esta pesquisa, foram utilizados não somente um item, mas um conjunto de setenta itens presentes com movimentação em noventa e oito centros de distribuição e que os melhores resultados foram obtidos quando também se agrupou os sub centros de distribuição para depois aplicar o modelo de previsão.

Yue *et al.* (2010) defenderam um modelo baseado em SVM para previsão em um varejista cervejeiro chinês, denominado *Selective and Heterogeneous SVM Ensemble for Demand Forecasting* (SHEnSVM), no qual diferentes modelos de SVM são treinados por diferentes amostras gerados por um algoritmo *Bootstrap* com diferentes parâmetros para garantir a diversidade de opções (EFRON, 1993). Em seguida, um algoritmo genético é aplicado para identificação do melhor cenário e parâmetros. Por fim, o SHEnSVM é utilizado com estes parâmetros para a previsão de demanda semanal. Conforme os autores, este modelo apresenta uma forte habilidade de generalização e uma melhor capacidade de previsão com relação às redes neurais para amostra de dados pequenos. Para implementação do modelo, o estudo utilizou sete atributos: preço da semana, fator sazonal, promoção, o efeito na demanda quando outras categorias estão em promoção, aumento da demanda quando em promoção, demanda das últimas quatro semanas e demanda da última semana. Como resultado do modelo para uma marca, o MAPE obtido do modelo proposto (SHEnSVM) e um modelo SVM simples foi de 2,26% e 2,73%,

respectivamente, indicando que o modelo proposto pode ser implementado para redução do nível de estoque e melhoria na política de ressuprimento.

Já Cao e Wu (2016) apresentaram uma proposta de um modelo híbrido de *Support Vector Regression* (SFOASVR) com um algoritmo de otimização *Fruit Fly* e ajuste dos parâmetros sazonais para previsão de demanda no consumo de energia mensal na China e no mercado varejista de energia dos Estados Unidos. Os autores ressaltam que o mercado de energia possui características complexas não lineares e uma forte tendência sazonal. Além disso, os autores ainda compararam a previsibilidade com outros algoritmos como o ARIMA sazonal (SARIMA) e BPNN através do MAPE. Os resultados mostraram que, para o mercado chinês, os valores de MAPE para o SFOASVR, SARIMA e BPNN, foram: 2,638%, 4,315% e 5,462%. Já os resultados referentes ao mercado norte americano foram: 2,236%, 3,278% e 1,592%. Os autores concluíram que o modelo proposto é superior com relação ao modelo linear SARIMA pelo fato dos dados se apresentarem não lineares. Já com relação ao modelo de redes neurais, o BPNN se mostra superior nos dados extensos, como o utilizado do mercado norte americano, mas quando os dados são pequenos, como os utilizados da China, o modelo SFOASVR se mostra superior.

Chen e Lu (2017) também adotaram em seus modelos um algoritmo baseado em SVR, além de ELM em combinação com métodos de clusterização para previsão de vendas em um varejista de eletrônicos. Primeiramente os dados são custerizados com padrões similares em diferentes grupos. Posteriormente o treinamento dos modelos de SVR e ELM são treinados para obtenção da previsão. Para a clusterização, são adotados três algoritmos diferentes, *self-organizing map* (SOM), *growing hierarchical self-organizing map* (GHSOM) e *K-means*, que em combinação com SVR e ELM, foi

realizado a comparação entre os seis possíveis modelos em dados bi-semanais para três grupos de produtos diferentes.

Os resultados médios das seis possíveis combinações dos modelos obtiveram os seguintes valores de MAPE: GHSOM-ELM: 8,72%, K-ELM: 15,03%, SOM-ELM: 15,26%, GHSOM-SVR: 13,11%, K-SVR: 20,10%, SOM-SVR: 21,93%.

Na pesquisa de Aras *et al.* (2017), a previsão de demanda semanal de um varejista multinacional de móveis localizado na Turquia é explorada através da comparação de quatro modelos preditivos, ARIMA, *Autoregressive Fractionally Integrated Moving Average* (ARFIMA), redes neurais artificiais (ANN) e *Adaptative Network-based Fuzzy Inference System* (ANFIS). O modelo ARFIMA, por trazer uma dependência de longo prazo com os preditores, ele se mostra, em alguns casos, resultados melhores quando é utilizado preditores financeiros, macroeconômicos e históricos de vendas agregados, (AYE, 2015; BHARDWAJ, 2006). Já o modelo ANFIS é uma combinação de rede neural, que possui a habilidade de identificação de relações complexas entre os preditores e a previsão, com sistemas *fuzzy*, que se baseiam na habilidade humana de inferência em casos de informações imprecisas.

Para a comparação são utilizados dez conjuntos de dados ao longo de 174 semanas, que representam diferentes produtos representativos da venda do varejista que apresentam diferentes padrões de vendas e que possuam uma dificuldade de previsão, conforme os especialistas da organização. A comparação revelou que nenhum modelo de sobressaiu em relação aos demais para todos dos dados analisados, sendo recomendada a combinação de mais de um modelo através da média, mediana, média desconsiderando os valores das extremidades ou médias considerando o MSE. O estudo, por fim, recomenda a adoção de outras redes neurais, juntamente com a adoção de variáveis

externas à organização como posicionamento da marca e fatia de mercado, além de outros que possam contribuir para a previsão. Nos resultados, o modelo GHSOM-ELM se mostrou a melhor performance com relação aos demais cinco modelos utilizados.

#### 2.4.2 ARTIGOS COM ATRIBUTOS INTERNOS E EXTERNOS

Di Pillo *et al.* (2016) definem no objetivo de pesquisa a previsão de demanda de uma *commodity* com grande volume de vendas em duas lojas físicas de um varejista, em um cenário de promoções em que a demanda se torna altamente volátil e com um comportamento não linear.

Com isso, é proposto um modelo baseado em *Support Vector Machine* (SVM), com treinamento em dados contendo nove atributos de calendário (mês, dia do mês e os sete dias da semana representados de forma booleana) e quatro atributos internos (item na promoção, horas de loja aberta, preço diário, quantidade de cupons emitidos no dia), sendo o último o valor a ser previsto pelo modelo. Os dados foram coletados em duas lojas da mesma cadeia de distribuição, porém com volumes de vendas distintos, sendo a loja 2 que possui o maior volume de vendas. Como resultado do MAPE, o modelo apresentou 4,5% e 5,4% para as lojas 1 e 2, respectivamente.

Com o objetivo de melhorar a logística em casos de emergência, Xue *et al.* (2018) propõem um modelo que combina um algoritmo baseado em *Neighborhood Rough Set* (NRS), que lida com atributos numéricos de forma a reduzir a dimensão da amostra, sem perder informações relevantes, juntamente com um algoritmo SVM, que lida com a não linearidade dos dados e dados não completos, com os parâmetros otimizados por um algoritmo genético (GA). Com isto, os dados foram submetidos a três modelos NRS-

SVM-GA, SVM e BPNN, que apresentaram o MAPE, com os respectivos valores: 0,63%, 7,67% e 14,20%. Portanto, o modelo proposto (NRS-SVM-GA) se mostrou eficiente na previsão de cenários de emergência, por apresentar o menor MAPE que os outros dois modelos citados, que conforme os autores, se adaptou melhor devido aos dados serem não lineares e com características de aleatoriedade.

Para o mercado de consumo de energia, em que os preços variam conforme a demanda, é muito importante a previsão de preço do kWh diário para planejamento da conta de energia do usuário. Neste sentido, Tee *et al.* (2017) propõem um modelo baseado em ELM, que considera dados do mercado como consumo diário e condições meteorológicas, para previsão do preço de energia. Na comparação de um modelo de média móvel com o ELM proposto, as medições do MAPE obtidas foram de 1,035% e 1,025%, respectivamente, indicando que a adoção de múltiplos dados torna o resultado mais preciso com relação à média móvel, que utiliza somente os dados históricos para previsão.

De forma similar, as vendas de diversos produtos perecíveis presentes em uma padaria também possuem demanda volátil, podendo acarretar custos com superprodução ou perda de vendas. Sendo assim, Huber *et al.* (2019) exploraram os dados de uma rede alemã de padarias, para obter uma previsão de demanda através de dois modelos orientados por dados: *Multi Layer Perceptron* (MLP) e Regressão Quantílica (RQ). A abordagem foi integrar a previsão de demanda de uma rede neural MLP e *Decision Trees* (DT) com um otimizador baseado em RQ, de forma a tornar os dois modelos aptos para tratarem sazonalidades e relações não lineares presentes nos dados.

Os dados utilizados nos modelos podem ser divididos entre internos e externos à organização. Os dados internos são: tipos de loja, categoria do produto, tempo de loja

aberta, vendas represadas, mediana de vendas, informações de promoções e localização da loja. Já os dados externos são: tipo de estabelecimento próximo às lojas (concorrentes, supermercados, açougues, redes de *fast food*, mecânicas, igrejas, clínicas médicas, *playgrounds*, parques, escolas e universidades), dados de calendário (dia do ano, mês, dia do mês, dia da semana, feriados, finais de semana, datas comemorativas e indicativos de datas especiais como ponto facultativo escolares), temperatura e exposição solar diários.

As melhores performances obtidas pelos modelos propostos MLP e DT, em comparação com SARIMA e *Seasonal Moving Average* (SMA), medida através do MAPE foram 27,87%, 28,34%, 29,71% e 30,15%, respectivamente. Com isso, os autores concluíram que a adoção de métodos de ML se torna viável caso haja dados suficientes. Neste caso havia pelo menos 150 linhas de informações. Caso contrário, métodos mais simples como SARIMA e SMA se tornam alternativas mais atraentes.

Adicionalmente, Caglayan *et al.* (2020) também utiliza múltiplos dados no segmento de moda como descontos, quantidade de clientes, mídias especiais, temperatura, localização de loja e histórico de vendas para treinamento do modelo proposto de redes neurais artificiais (ANN) a fim de obter uma previsão de demanda para os produtos de um varejista de sapatos esportivos localizado na Turquia. Para a definição do tipo da rede neural, quantidade de camadas ocultas, quantidade de neurônios nas camadas ocultas, funções de ativação e transferência foram obtidas através de múltiplas tentativas para chegar no melhor resultado do MAPE, igual a 8,79%. Não foi realizada uma comparação do modelo proposto com outros modelos. Os autores fazem uma ressalva de que o modelo proposto poderá mudar seus parâmetros e estrutura para previsão de outros segmentos de produtos nas lojas que possuam diferentes índices de sazonalidade.

Para o mercado de produtos de alto valor agregado no varejo não alimentar de lojas físicas, que possuem uma grande variação na demanda, Qu *et al.* (2017) adotaram um modelo de *Random Forest* (RF) para predição da demanda semanal destes itens, utilizando para isto informações de precificação, feriados, descontos, quantidade de estoque e fatores regionais (temperatura, localização de loja, preço de combustíveis e taxa de desemprego). Em seguida, é realizada uma otimização para definição de uma política de melhor preço através de um método heurístico. Os melhores resultados do MAPE obtido para as previsões do modelo proposto RF e Regressão Multivariada foram de 23,6% e 37%, respectivamente. Os autores sugerem um estudo futuro que agregue informações do mercado *online* e *offline*, pois este cenário de omnicanalidade está cada vez mais presente e relevante para os consumidores tomarem suas decisões de compras.

Um modelo também baseado em *Random Forest* (RF) é adotado por Glaeser *et al.* (2019) para definir o melhor local, data e horário para coleta de produtos comprados *online* em que a coleta é realizada em um veículo, que pode estar em qualquer ponto da cidade. Esta definição é baseada conforme a previsão em locais potenciais, na qual são combinados dados demográficos e econômicos (população total, população feminina, população masculina, população com nível médio concluído, quantidade de casas, quantidade de casas com menores de idade, quantidade de casas com renda acima de US\$75.000,00 anuais, média de idade, média no valor da residência, renda média e renda mediana, quantidade de empregados por milha quadrada), de localização (identificação da localização, nome da localização, tipo do local, ponto próprio, centros de distribuição que atendem o local, latitude e longitude), histórico de vendas (data, dia da semana, quantidade vendida, quantidade de descontos) e operacionais. Desta forma, o modelo propõe evitar canibalização de pontos de coleta e aumentar as vendas através de um melhor nível de serviço.

O modelo de previsão proposto por Glaeser *et al.* (2019) foi comparado com outros três modelos, *Forward Selection Regression*, SVM e Redes Neurais, obtendo resultados de MAPE, respectivamente, iguais a 19%, 50%, 48% e 99%. Esta pesquisa mostra como é possível utilizar dados públicos externos e de acesso a todos, em combinação com os dados internos de vendas e operação para obter uma previsão de demanda que seja relevante para tomada de decisões táticas e estratégicas em uma empresa.

Guo *et al.* (2012) utilizaram o modelo proposto por Wong e Guo (2010) para incluir variáveis internas (preço de venda original, quantidade de compra, data de lançamento, tempo de duração do produto, 4 volumes de vendas anteriores) e externas (dados meteorológicos e 5 indicadores econômicos: PIB, CPI - *Consumer Price Index*, PPI - *Producer Price Index*, índice total de vendas e índice de preço da categoria) para a previsão de demanda em um varejista de moda chinês. Apesar de não apresentar os valores do MAPE para os modelos adotados, é indicado que o modelo proposto obtém uma quantidade superior em relação aos demais com relação a quantidade de vezes em que ele obteve um MAPE menor que os demais, utilizando a mesma base de dados.

O quadro 2 resume os artigos com as informações do algoritmo adotado, os atributos internos e externos utilizados, o segmento do conjunto de dados e o resultado dos principais indicadores calculados, respectivamente.

**Quadro 2.** Resumo dos artigos presentes na revisão literária no período de 2010 a 2020.

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Wong W.K., Guo Z.X.	Int J Prod Econ	A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm	2010	Neural Network; Heuristic Fine-tuning Process	Histórico de vendas		Moda	<p>Previsão Mensal (NN):                      Categoria 1 = 19,3%                      Categoria 2 = 25,6%                      Categoria 3 = 37,0%                      Categoria 4 = 58,1%</p> <p>Previsão Trimestral (NN):                      Categoria 1 = 11,9%                      Categoria 2 = 17,2%                      Categoria 3 = 15,1%                      Categoria 4 = 16,2%</p> <p>Previsão Anual (NN):                      Categoria 1 = 6,5%                      Categoria 2 = 11,1%                      Categoria 3 = 7,0%                      Categoria 4 = 24,9%</p>	<p>Previsão Mensal (ARIMA):                      Categoria 1 = 31,8%                      Categoria 2 = 38,1%                      Categoria 3 = 35,5%                      Categoria 4 = 6665,4%</p> <p>Previsão Trimestral (ARIMA):                      Categoria 1 = 15,3%                      Categoria 2 = 27,2%                      Categoria 3 = 20,2%                      Categoria 4 = 40,5%</p> <p>Previsão Anual (ARIMA):                      Dados insuficientes para estimar o modelo</p>
Yue L., Zhenjiang L., Yafeng Y., Zaixia T., Junjun G., Bofeng Z.	Proc. - IEEE Int. Conf. Comput. Inf. Technol., CIT, IEEE Int. Conf. Embedded Softw. Syst., ICES, ScalCom	Selective and Heterogeneous SVM Ensemble for demand forecasting	2010	Support vector machine; genetic algorithm	Preço da semana vigente, fator sazonal, promoções de vendas na semana, efeito na quantidade vendida devido a promoções em outras categorias na mesma semana, incremento na demanda na categoria devido a promoções vigentes, média de vendas das últimas quatro semanas, quantidade de vendas na última semana		Varejo de bebidas	<p>Previsão Semanal (SHEnSVM)                      2,26%</p>	<p>Previsão Semanal (SVM)                      2,73%</p>

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Guo Z.X., Li M., Wong W.K.	IEEE Int. Conf. Ind. Informatics (INDIN)	Intelligent multivariate sales forecasting using wrapper approach and neural networks	2012	harmony search- wrapper-based variable selection (HWVS); multivariate intelligent forecaster (MIF) module; Neural Networks	Preço de venda original, quantidade de compra, data de lançamento, tempo de duração do produto, 4 volumes de vendas anteriores	dados meteorológicos e 5 indicadores econômicos (PIB, CPI - Consumer Price Index, PPI - Producer Price Index, índice total de vendas e índice de preço)	Varejo generalista	-	-
Wong W.K., Guo Z.X.	Optim. Decis. Mak. in the Appar. Supply Chain Using Artif. Intell. (AI): From Prod. to Retail	Intelligent sales forecasting for fashion retailing using harmony search algorithms and extreme learning machines	2013	ANN; Extreme Learning Machine; procedimento heurístico	Histórico de vendas		Moda	<p>Previsão Mensal (NN): Categoria 1 = 19,3% Categoria 2 = 25,6% Categoria 3 = 37,0% Categoria 4 = 58,1%</p> <p>Previsão Trimestral (NN): Categoria 1 = 11,9% Categoria 2 = 17,2% Categoria 3 = 15,1% Categoria 4 = 16,2%</p> <p>Previsão Anual (NN): Categoria 1 = 6,5% Categoria 2 = 11,1% Categoria 3 = 7,0% Categoria 4 = 24,9%</p>	<p>Previsão Mensal (ARIMA): Categoria 1 = 31,8% Categoria 2 = 38,1% Categoria 3 = 35,5% Categoria 4 = 6665,4%</p> <p>Previsão Trimestral (ARIMA): Categoria 1 = 15,3% Categoria 2 = 27,2% Categoria 3 = 20,2% Categoria 4 = 40,5%</p> <p>Previsão Anual (ARIMA): Dados insuficientes para estimar o modelo</p>

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Islek I., Oguducu S.G.	IEEE Int Symp Ind Electron	A retail demand forecasting model based on data mining techniques	2015	Bayesian Network Algorithm	Média móvel de vendas, atributos de centros de distribuição (localização, quantidade de subcentro de distribuição, quantidade de veículos, quantidade de vendas semanais dos produtos), área de venda, quantidade de empregados, quantidade de clientes, atributos relacionados a produtos (quantidade de vendas e período de vendas)		Varejo generalista	Previsão Semanal (Média Móvel + Rede Bayesiana) Sem clusterização: 49% Clusterização dos centros de distribuição: 24% Clusterização dos sub centros de distribuição: 17%	Previsão Semanal (Média Móvel) Sem clusterização: 129%
Cao G., Wu L.	Energy	Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting	2016	Support Vector Regression (SVR); Fruit Fly Optimization Algorithm	Histórico de vendas		Energia	Modelo SFOASVR Consumo China = 2,638% Varejo EUA = 2,236%	Modelo SARIMA Consumo China = 4,315% Varejo EUA = 3,278%  Modelo BPNN Consumo China = 5,462% Varejo EUA = 1,592
da Veiga C.P., da Veiga C.R.P., Puchalski W., Coelho L.S., Tortato U.	J. Retail. Consum. Serv.	Demand forecasting based on natural computing approaches applied to the foodstuff retail segment	2016	Wavelets Neural Networks; Takagi-Sugeno Fuzzy System	Histórico de vendas		Alimentos perecíveis	Modelo TS Categoria A = 5,73% Categoria B = 5,72% Categoria C = 5,51%  Modelo WNN Categoria A = 5,46% Categoria B = 4,88% Categoria C = 4,63%	Modelo ARIMA Categoria A = 6,52% Categoria B = 7,23% Categoria C = 8,57%

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Di Pillo, G; Latorre, V; Lucidi, S; Procacci, E	4Or-A Quarterly Journal Of Operations Research	An application of support vector machines to sales forecasting under promotions	2016	Support Vector Machine	4 atributos específicos do problema (promoção, horas de funcionamento de loja, preço diário, número de cupons emitidos)	9 atributos de calendário (mês, dia do mês, dia da semana, este é representado por 7 atributos booleanos exclusivos que levam em consideração comportamento e hábitos dos consumidores, por exemplo, domingo é esperado uma venda maior com relação aos demais dias da semana)	Varejo de commodities	Previsão (SVM) Loja 1 = 4,5% Loja 2 = 5,4%	
Aras S., Deveci Kocakoç I., Polat C.	J. Bus. Econ. Manage.	Comparative study on retail sales forecasting between single and combination methods	2017	Exponential Smoothing (ETS), ARIMA, ARFIMA, Artificial Neural Networks (ANN), Adaptive Network-based Fuzzy Inference System (ANFIS)	Histórico de vendas	Mobília	Modelo ANFIS Data 1 = 51,29% Data 2 = 14,66% Data 3 = 20,04% Data 4 = 33,37% Data 5 = 21,46% Data 6 = 18,72% Data 7 = 34,81% Data 8 = 19,47% Data 9 = 22,35% Data 10 = 17,20%	Modelo ARIMA   ANN   ARFIMA Data 1 = 51,29%   52,79%   53,54% Data 2 = 14,66%   14,76%   15,78% Data 3 = 20,04%   20,87%   21,15% Data 4 = 33,37%   37,00%   31,13% Data 5 = 21,46%   17,97%   20,31% Data 6 = 18,72%   17,66%   17,81% Data 7 = 34,81%   39,77%   34,59% Data 8 = 19,47%   20,10%   22,42% Data 9 = 22,35%   22,66%   19,99% Data 10 = 17,20%   20,16%   17,07%	

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo	
Chen, IF; Lu, CJ	Neural Computing & Applications	Sales forecasting by combining clustering and machine-learning techniques for computer retailing	2017	Three clustering techniques including self-organizing map (SOM), growing hierarchical self-organizing map (GHSOM), and K-means and two machine learning techniques including support vector regression (SVR) and extreme learning machine (ELM) are used in this study.	Histórico de vendas		Varejo de eletrônicos	Produto		
								Modelo	MAPE	
								1	GHSOM-SVR (6)	14,72%
								1	K-SVR (6)	25,44%
								1	SOM-SVR (6)	24,56%
								1	GHSOM-ELM (5)	7,42%
								1	K-ELM (5)	17,83%
								1	SOM-ELM (6)	15,55%
								2	GHSOM-SVR (5)	13,97%
								2	K-SVR (6)	17,78%
								2	SOM-SVR (3)	21,36%
								2	GHSOM-ELM (4)	10,43%
								2	K-ELM (4)	15,59%
								2	SOM-ELM (5)	18,75%
3	GHSOM-SVR (5)	10,65%								
3	K-SVR (6)	17,09%								
3	SOM-SVR (6)	19,87%								
3	GHSOM-ELM (5)	8,31%								
3	K-ELM (5)	11,67%								
3	SOM-ELM (5)	11,48%								
Qu T., Zhang J.H., Chan F.T.S., Srivastava R.S., Tiwari M.K., Park W.-Y.	Comput Ind Eng	Demand prediction and price optimization for semi-luxury supermarket segment	2017	Regression Tree; Random Forest	preço, descontos, inventário, marca, design, forma, cor e outros parâmetros	data de feriados, temperatura, preços de combustíveis, taxa de desemprego	Varejo de commodities de alto valor agregado	RF: 23,6%	Regressão Multivariada: 37%	

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Tee J.E.C., Teo T.T., Logenthiran T., Woo W.L., Abidi K.	IEEE Reg 10 Annu Int Conf Proc TENCON	Day-ahead forecasting of wholesale electricity pricing using extreme learning machine	2017	Artificial Neural Networks; Extreme Learning Machines (ELM)	histórico de vendas e precificação	dados meteorológicos	Energia	ELM: 0,25%	Média Móvel: 1,035%
Loureiro A.L.D., Miguéis V.L., da Silva L.F.M.	Decis Support Syst	Exploring the use of deep neural networks for sales forecasting in fashion retail	2018	Deep Learning, Deep Neural Networks	Características físicas dos produtos, preço e atributos (família, subfamília, tipo de cor, moda, segmento, tipo de loja, tamanho, e nível de expectativa)		Moda	Previsão por Temporada DNN: 37,8%	Previsão por Temporada DT: 37,7% RF: 34,5% SVR: 39,3% ANN: 38,7% RL: 45,1%
Xue H., Jiang C., Cai B., Yuan Y.	Proc. Chin. Control Decis. Conf., CCDC	Research on demand forecasting of retail supply chain emergency logistics based on NRS-GA-SVM	2018	NRS-GA-SVM Algorithm; Neighborhood Rough Set (NRS); Support Vector Machines (SVM)	Área da loja, Histórico de vendas, Demanda diária de materiais da loja localização e quantidade de lojas	Variedade dos desastres, grau dos desastres, grau de obstrução na divulgação das informações do desastre, consumo per capita dos suprimentos emergenciais fornecidos pelo governo, dano a residências, dano aos suprimentos, duração do desastre, tipo da localização do desastre, material emergencial original, clima ao redor das lojas afetadas, status dos materiais alternativos afetados, coeficiente de lojas afetadas, coeficiente de desastres naturais, tráfego ao redor das lojas e nível de recuperação dos estabelecimentos ao redor das lojas	Logística de emergência	NRS-SVM-GA: 0,63%	SVM: 7,67% BPNN: 14,20%

Autores	Periódico	Título	Ano	Algoritmo	Atributos Internos	Atributos Externos	Segmento	Resultado MAPE (Modelo Proposto)	Resultado MAPE Comparativo outro modelo
Glaeser C.K., Fisher M., Su X.	Manuf. Serv. Oper. Manage.	Optimal retail location: Empirical methodology and application to practice	2019	Random Forests	ID da localização, nome da localização, tipo de localização (comercial, escola, etc.), total de vendas, total de cupons de desconto, posse do imóvel, ID do centro de distribuição que atende, data, dia da semana, latitude, longitude, e disponibilidade de entrega	Total da população, gênero, grau de escolaridade, quantidade de residências, quantidade de residências com menores de idade, quantidade de residências com renda superior a US\$75.000/ano, média de idade, valor médio da residência, renda média, renda mediana, empregados por milha quadrada	Varejo Online	RF: 19%	Forward Selection Regression: 50% SVM: 48% Neural network: 99%
Huber J., Müller S., Fleischmann M., Stuckenschmidt H.	Eur J Oper Res	A data-driven newsvendor problem: From data to decision	2019	Decision Trees; ANN; Quantile Regression	Histórico de vendas, classe de loja, categoria do produto, vendas atrasadas, mediana móvel de vendas, informação binária de promoções (S/N), tipo de localização (cidade, vila, subúrbio)	Aberturas (dias, duração em horas), dia do ano, mês, dia do mês, dia da semana, feriado público, tipo do dia, data de comemoração festiva, dia não útil, indicadores de dias especiais, feriados escolares, temperatura (mínima, média, máxima) e cobertura de nuvens por dia, proximidades da loja (padarias, açougues, lojas de conveniência, quiosques, lojas de fast-food, mecânicas, etc.), pontos de saúde e conforto (hospitais, consultórios, igrejas), pontos de lazer (playground, locais esportivos, parques), pontos de educação (creches, escolas, universidades)	Padaria	MLP: 27,87% DT: 28,34%	SARIMA: 29,71% SMA: 30,15%
Caglayan N., Satoglu S.I., Kapukaya E.N.	Adv. Intell. Sys. Comput.	Sales forecasting by artificial neural networks for the apparel retail chain stores	2020	Artificial Neural Networks	Histórico de vendas, faturamento, descontos, número de clientes e localização das lojas	Temperatura, feriados	Moda	ANN: 8,79%	

## 2.5 ANÁLISE DE DADOS

Conforme Breiman (2001), existem duas abordagens para modelagem dos dados para que se possa chegar a conclusões de predição. A primeira é regida por modelagem estocástica dos dados, ou seja, para dados pertencentes a um processo aleatoriamente determinado teremos um modelo equacional (*e.g.* Regressão Linear Múltipla e Logística) para predição da variável dependente  $Y$  conforme os dados de entrada  $X$ . Para isto, o pesquisador deverá assumir algumas hipóteses sobre a natureza do processo. Caso o resultado do modelo seja satisfatório, com uma boa aproximação dos dados reais, então é possível realizar inferências sobre a natureza do processo. Nesta abordagem, as conclusões são sobre o modelo proposto e não sobre o processo real (HAIR, 2018).

A segunda abordagem feita através de mineração de dados são os modelos algorítmicos que tratam os dados de entrada  $X$  em uma caixa preta (*e.g.* RNA, *Deep Learning*), porém oferece os sinais de saída  $Y$  de maneira preditiva com alta acuracidade tanto em bases dados de pequeno e de grande porte. Para Hair *et al.* (2018), a explicação do processo é uma etapa secundária para esta abordagem, sendo a primária a capacidade de predição. Uma vantagem que Breiman (2001) defende para o uso da abordagem algorítmica é a possibilidade de sua implementação em dados altamente complexos e com grande volume de informações sem a necessidade de especificação de um modelo que requeira um grande esforço e, em alguns casos, nem seria possível. A mineração de dados também é conhecida por *Machine Learning* e Inteligência Artificial, conforme Hair *et al.* (2018).

Hair *et al.* (2018) defende que não existe uma abordagem única, mas que é importante termos ciência das ferramentas disponíveis para análise dos dados e que em cada situação

haverá pontos positivos e negativos de cada uma que, ao final, são complementares para o pesquisador.

### 3 ABORDAGEM METODOLÓGICA

Em um cenário de omnicanalidade, o varejo ganha complexidade nas informações disponíveis para tomada de decisão e definição de políticas para o atendimento da demanda em múltiplos pontos de entrega. Estes dados são provenientes de fontes internas e externas da organização, com natureza diferentes (dados regulares e não regulares, lineares e não lineares, numéricos, textos, fotos, indicadores econômicos, dados estatísticos públicos, dados meteorológicos, etc.) e um grande volume de dados que não tornaria prático e eficiente a utilização de diversas fontes para uma previsão de demanda por canal e evolução do nível de serviço.

Com isso, a adoção de algoritmos de inteligência artificial, que sejam capazes de avaliar qual dado é mais relevante na previsão de demanda e em um tempo de resposta ágil, faz sentido e se mostra como um fator que impacta na satisfação dos clientes inseridos neste contexto e redução de despesas operacionais.

Portanto, há uma oportunidade na comparação dos algoritmos observados na literatura com o método proposto que utiliza um algoritmo de *Hierarchical Clustering*, seguido por uma Rede Neural Artificial para previsão de demanda no contexto de omnicanalidade no varejo com adoção de atributos internos e externos, conforme identificado por Melacini *et al.* (2018), que ainda estão na fase inicial de desenvolvimento e se mostra uma tendência de pesquisa e de adoção obrigatória a curto prazo pelos varejistas. Ainda não foi identificado na literatura um estudo com esta abrangência.

### 3.1 DELIMITAÇÃO DA PESQUISA

A pesquisa se delimita a um estudo de caso operacional em que há a aplicação de dois algoritmos de inteligência artificial (*Hierarchical Clustering* e Rede Neural Artificial) nos dados do objeto de estudo. É verificada a performance do modelo através de métricas de assertividade e comparado os resultados com os modelos presentes na literatura para previsão de demanda de um varejo em ambiente de omnicanalidade.

A limitação do método está na adoção de uma abordagem conjunta de dois algoritmos de inteligência artificial, treinados por um banco de dados em um período de dois anos de uma loja física no setor de varejo alimentar, através das transações de vendas *online* e *offline*, em presença de omnicanalidade. Os dados internos se restringem às características dos produtos (preço, quantidade, tipo de produto) e não sobre informações dos clientes (perfil sócio econômico, idade, gênero, etc.), o que permitiria ao estudo uma análise mais detalhada entre o comprador e vendedor.

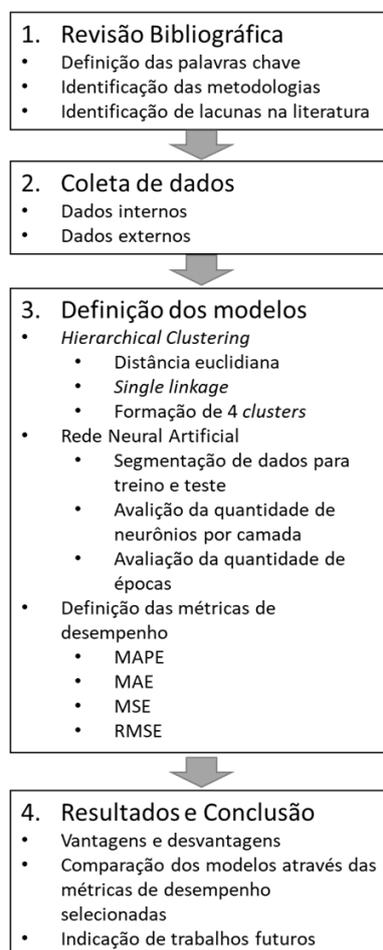
### 3.2 OBJETO DE PESQUISA

O objeto de pesquisa é uma empresa de varejo de bens de consumo que está inserida em um ambiente de omnicanalidade, isto é, oferece ao consumidor pelo menos dois canais de venda integradas com possibilidades de múltiplos pontos de interação (loja física, aplicativos móveis e internet) localizada na cidade de Goiânia-GO. O estabelecimento é de médio porte (1800 m<sup>2</sup> de piso de loja) com 145 funcionários e 11 checkouts. Oferece um mix de 14000 produtos alimentares e itens de bazar. Os dados disponibilizados foram registros de vendas entre o período de 2019 a 2020 do seu maior fornecedor de bebidas, conforme as restrições de disponibilização dos dados pelo próprio estabelecimento,

totalizando transações de venda de 138 produtos distintos do estabelecimento (águas, sucos, chás, isotônicos, refrigerantes e cervejas) com representatividade de 3% do total do faturamento da loja. O software utilizado para extração dos dados foi o Winthor/TOTVS através de uma instrução *Structured Query Language* (SQL).

### 3.3 MÉTODO DE PESQUISA

A metodologia adotada seguiu a ordem apresentada na figura 5.



**Figura 5.** Metodologia de pesquisa para proposição e avaliação do modelo de previsão de demanda.

**Fonte:** Elaborado pelo autor (2021)

Foi realizado uma revisão bibliográfica na base SCOPUS com as *keywords* ((TITLE-ABS-KEY(learning) OR TITLE-ABS-KEY(neural) OR TITLE-ABS-KEY(machine)) AND (TITLE-ABS-KEY(forecast) OR TITLE-ABS-KEY(demand)))

AND TITLE-ABS-KEY(retail)) e no Web of Science com as *keywords* ((ALL=(learning) OR ALL=(neural) OR ALL=(machine)) AND (ALL=(forecast) OR ALL=(demand)) AND ALL=(retail)) a com o período de publicação entre 2010 a 2020, com o levantamento de 124 artigos não duplicados nestas duas bases e revisados por pares.

A seleção dos artigos foi realizada através da leitura do título, resumo e conclusão, separando assim, 43 artigos que tem como foco de pesquisa a previsão de demanda no varejo utilizando algoritmos de inteligência artificial. Foram identificadas quais metodologias são mais utilizadas, o setor de bens e serviços, os algoritmos adotados, os atributos internos e externos utilizados, e os resultados atingidos em cada abordagem. Posteriormente, foram identificadas quais lacunas estavam presentes na literatura com relação à adoção de modelos de inteligência artificial para auxiliar na tomada de decisões na cadeia de suprimentos, com base na previsão de demanda (LAGORIO, 2016).

Em seguida, foram coletados dados de um varejo em cenário de omnicanalidade de um período de 24 meses entre janeiro de 2019 e dezembro de 2020, com as seguintes informações:

- Data;
- Identificação do Produto;
- Canal (loja física ou pedido *online*);
- Preço unitário;
- Quantidade de venda.

Ao total foram coletados 42335 registros de informação, sendo que cada registro corresponde a venda diária de cada produto. Os dados foram extraídos por uma consulta no banco de dados através de um código SQL.

A venda de itens em um varejo é diretamente impactada pelos eventos de calendário como feriados, finais de semana e datas comemorativas (FILDES, 2019). A consulta foi restrita a um fornecedor de bebidas, cujos dados tendem a sofrer mais impacto com relação aos atributos externos que foram considerados para o modelo:

- Final de semana ou feriado (sim/não);
- Precipitação diária (em mm);
- Temperatura máxima diária;
- Preço promocional (sim/não);
- Bebida alcoólica (sim/não);
- Teor alcoólico (% ABV – *alcohol by volume*, definido como o % de etanol presente em 100 mL de solução à 20°C);
- Medição do amargor da bebida (IBU – *International Bittering Unit*);

Inicialmente, os dados foram tratados em uma planilha eletrônica (MS Excel), assim os dados foram agregados em 3807 registros por:

- Data;
- Canal de venda (*online/offline*);
- Final de semana ou feriado (sim/não);
- Preço promocional (sim/não);
- Bebida alcoólica (sim/não);
- Teor alcoólico (% ABV);
- Medição do amargor da bebida (IBU);
- Precipitação (mm);
- Temperatura Máxima (°C);
- Quantidade de Venda (em unidades).

Na tabela três é apresentada a correlação entre as variáveis “Média de %ABV”, “Média de IBU” e “Alcoólico”, pois todas estão relacionadas às bebidas alcoólicas. As demais variáveis não apresentam uma correlação linear forte, porém, o modelo *data driven* proposto através de algoritmos de ML buscará também relações não lineares entre os dados apresentados durante o processo de aprendizagem de máquina. Isso será verificado e avaliado através das métricas de desempenho.

**Tabela 3.** Tabela de correlação entre variáveis

	ALCOOL ICO	Média de %ABV	VENDA_O NLINE	FINALSEMANA FERIADO	PROMO	IBU	PRECIPI TAÇÃO	TEMP_ MAX	QTD_ VEND A
ALCOOLICO	1								
Média de %ABV	0,99341	1							
VENDA_ONLINE	-0,0265	-0,03418	1						
FINALSEMANA FERIADO	0,00176	0,00516	-0,02391	1					
PROMO	0,02673	0,01309	-0,16749	-0,00390	1				
Média de IBU	0,94372	0,92933	-0,04729	0,00469	0,00426	1			
PRECIPITAÇÃO	-0,0127	-0,01374	-0,01047	0,02101	0,00413	-0,01114	1		
TEMP_MAX	0,00488	-0,00097	-0,06792	-0,03428	-0,02528	0,01707	-0,2200	1	
QTD_VENDA	0,33071	0,34130	-0,36178	0,18182	-0,03621	0,33089	-0,0213	0,03292	1

As variáveis binárias descrevem as seguintes características dos produtos:

- Alcoólico: 0 caso não seja uma bebida alcoólica e 1 caso seja alcoólica;
- Venda Online: 0 para vendas *offline* e 1 para vendas *online*;
- Final Semana Feriado: 0 para vendas entre segunda-feira a quinta-feira e 1 para vendas entre sexta-feira a domingo e feriados;
- Promo: 0 para vendas regulares e 1 para vendas com preço promocional;

As demais variáveis não binárias descrevem as seguintes características:

- % ABV: indica o % de álcool por volume da bebida para indicar o seu teor alcoólico;

- IBU: indica o grau de amargor da bebida através da escala de *International Bitterness Units*;
- Precipitação: indica a quantidade de chuva em milímetro (mm). Uma pluviosidade de 1 milímetro equivale ao volume de 1 litro (L) de água de chuva que se acumulou sobre uma superfície de área igual a 1 metro quadrado.
- Temperatura Máxima: indica a temperatura máxima registrada no dia da venda efetuada.

Posteriormente, implementados ao modelo proposto através da linguagem Julia, sendo de código aberto e distribuído sob a licença do *Massachusetts Institute of Technology* (MIT) (BEZANSON, 2017).

Foi proposto um modelo baseado em um algoritmo de *Hierarchical Clustering* para agrupamento dos dados com características semelhantes, seguido de um segundo algoritmo baseado em Redes Neurais Artificiais, utilizado para previsão de demanda por canal de vendas com a segmentação dos dados na proporção de 70% para treinamento e 30% para teste de generalização do modelo. Por fim, foi comparado o resultado do modelo proposto com os modelos presentes na revisão de literatura. Com isso, foi realizada uma comparação das principais métricas de desempenho de cada modelo para a tomada de decisão suportada pela previsão de demanda de produtos no cenário apresentado e possível generalização para outros cenários e dados em estudos futuros.

### 3.4 INSTRUMENTOS DA PESQUISA

Foi adotada como ferramenta de pesquisa a linguagem de programação Julia, versão 1.4.2, e sua biblioteca Flux, dedicada a implementação de ML, para construção, implementação e avaliação do algoritmo de inteligência artificial nos dados coletados.

A escolha pela linguagem Julia se deu por ser de código aberto e por combinar a produtividade e facilidade de codificação presente no Python com a performance de execução da linguagem C em problemas de computações numéricas intensos como algoritmos de redes neurais artificiais (BEZANSON, 2017). Além disso, é possível integrar no código as bibliotecas presentes nas linguagens Python e R caso necessário, tornando assim a linguagem Julia muito flexível com relação as ferramentas disponíveis atualmente.

O *framework* Flux foi escolhido por sua simplicidade, facilidade de alterações nos parâmetros e *design* de modelos de ML, além de sua performance de execução computacional (INNES, 2018).

## 4 RESULTADOS

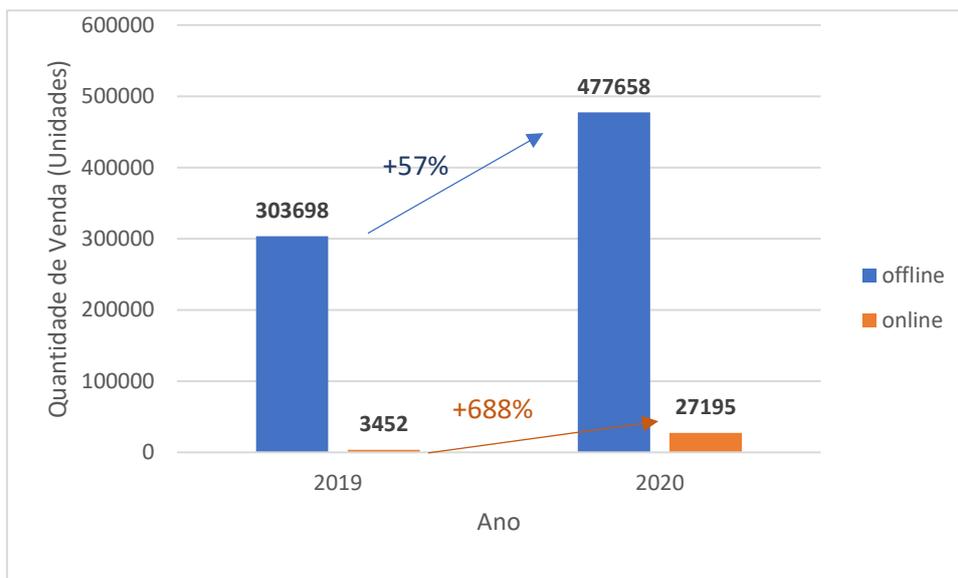
Os dados agregados apresentaram 3807 registros com a estatística descritiva apresentada pelo quadro três.

**Quadro 3.** Estatística descritiva dos dados.

	ALCOOLICO	VENDA_ONLINE	FINALSEMANA_FERIADO	PREÇO_PROMOCIONAL	QTD_VENDA	Média de %ABV	Média de IBU	PRECIPITAÇÃO	TEMP_MAX
Média	0,502	0,2635	0,4371	0,4331	213,2921	2,5576	6,4895	3,6022	30,976
Erro padrão	0,0081	0,0071	0,008	0,008	4,8884	0,0416	0,1058	0,1469	0,0471
Mediana	1	0	0	0	113	4,4	7	0	30,7
Modo	1	0	0	0	6	0	0	0	31
Desvio padrão	0,5001	0,4406	0,4961	0,4956	301,6186	2,5649	6,5253	9,0662	2,9076
Variância da amostra	0,2501	0,1941	0,2461	0,2456	90973,8022	6,5785	42,5794	82	8
Curtose	-2,001	-0,8462	-1,9366	-1,9282	29,2075	-1,9272	-1,4313	13,5625	0,6249
Assimetria	-0,0079	1,0743	0,2538	0,2699	3,9504	0,036	0,2782	3,4435	0,4163
Intervalo	1	1	1	1	4133	8	34,33333	66	19,1
Mínimo	0	0	0	0	0	0	0	0	21,4
Máximo	1	1	1	1	4133	8	34,3333	66	40,5
Soma	1911	1003	1664	1649	812003	9736,9	24705,6	13713,4	117925,7

Pela estatística de média as variáveis binárias “Alcoólico”, “Venda Online” e “Final de semana/feriado” indicam que as bebidas alcoólicas representam a metade dos registros. 50,2%, das vendas online representam 26,35% do total, sendo 43,71% aos finais de semana ou feriados. Foram considerados sexta-feira, sábado e domingo como dias de final de semana. Os registros apresentam 43,31% das transações quando os preços estão em valores promocionais, verificados através da média dos seus valores binários.

O crescimento na quantidade de venda no canal *online* no ano de 2020 com relação ao ano de 2019 foi de 688%, enquanto no canal *offline*, o crescimento foi de 57%, conforme observado pela figura 6. Indicando, assim, que o canal *online* está ganhando significativa relevância às vendas e decisões operacionais deste estabelecimento de varejo em omnicanalidade. Este crescimento mais forte no canal *online* pode estar relacionado à pandemia de COVID-19, em que foi observado o crescimento da digitalização em diversos setores da economia devido às restrições sociais.



**Figura 6.** Evolução de vendas de 2019 a 2020 nos canais *online* e *offline*.

**Fonte:** Elaborado pelo autor (2021)

#### 4.1 CLUSTERIZAÇÃO

Para a construção de um modelo de previsão de demanda, Chen et al. (2015) primeiramente agrupa os produtos com características semelhantes dos atributos ao longo do período. Portanto foi utilizada uma análise de *cluster* para identificar os produtos em grupos com maior similaridade, utilizando o software Orange, versão 3.27.1, que é uma plataforma de uso livre (licença GPLv3), com ferramentas de análise de dados, classificação, clusterização e previsão com interface gráfica, além de permitir a expansão das suas análises em linguagem Python. Seu desenvolvimento foi realizado pela Universidade de Ljubljana, Eslovênia e sua distribuição é realizada através do site <https://orangedatamining.com/>.

Seguindo Punj e Stewart (1983); Lee e Jung (2020), foi utilizado o procedimento de criar um modelo de análise baseado em *Hierarchical Clustering* (HC) com o cálculo

das distâncias com a métrica euclidiana através de *single linkage*, na qual, é considerada a distância mínima entre os membros dos dois *clusters*, para assim, identificar os *clusters* com os produtos que possuem maior similaridade. O dendrograma que representa a clusterização está apresentado pela figura 7.

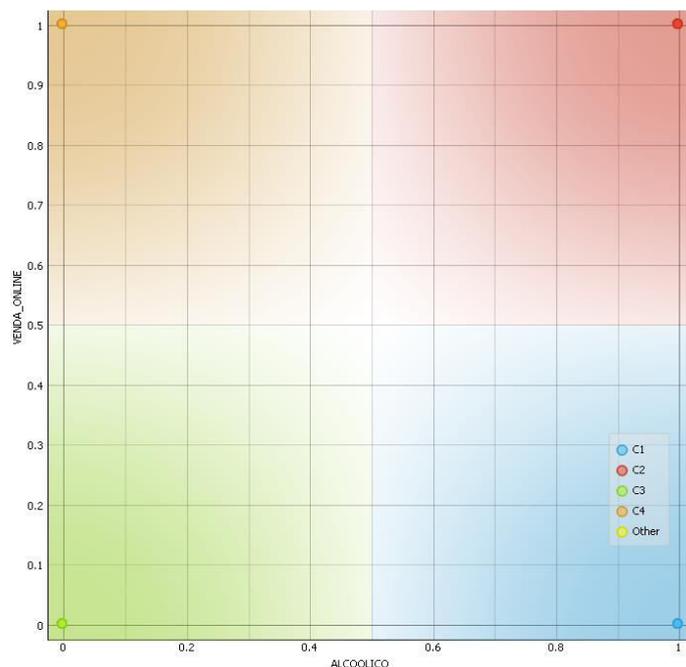
Pelo dendrograma, foi possível visualizar e segmentar os produtos em quatro *clusters*.



**Figura 7.** Agrupamento dos produtos em quatro *clusters*.

**Fonte:** Elaborado pelo autor (2021)

A divisão em quatro *clusters*, definida pelo algoritmo HC, apresentou a segregação bem definida entre vendas *online* e *offline* e bebidas alcoólicas e não alcoólicas em quatro quadrantes bem definidos e sem sobreposição de registros em mais de um *cluster*, conforme apresentado na figura 8, no qual as vendas *offline* foram representadas pelos *clusters* C1 e C3, com 1049 e 1375 registros ou 73,7% do total, enquanto que a venda *online* foram representados pelos *clusters* C2 e C4 com 479 e 520 registros ou 26,3% do total. Portanto será possível entender o comportamento dos modelos propostos conforme o canal de vendas e tipo de bebidas.



**Figura 8.** Representação dos quatro *clusters* nos quadrantes de venda *online*, *offline*, alcoólicos e não alcoólicos.

**Fonte:** Elaborado pelo autor (2021)

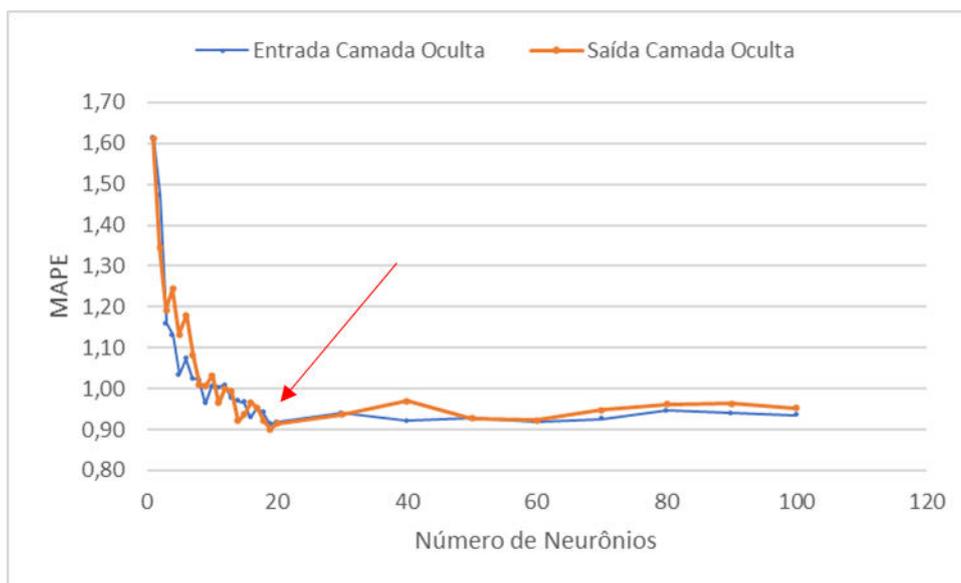
Em seguida, foi realizada uma segmentação dos dados na proporção de 70% para treinamento dos modelos e 30% para teste de precisão, conforme proposto por Livieris *et al.* (2019) em que foram testadas diversas proporções e em diferentes algoritmos e que esta proporção obteve uma melhor previsibilidade.

## 4.2 REDES NEURAIAS ARTIFICIAIS

Por fim, foi implementado através da biblioteca Flux desenvolvida por Innes *et al.* (2018), os dados no modelo de Redes Neurais Artificiais, sendo este, uma rede com duas camadas ocultas.

A definição da quantidade apropriada de neurônios nas camadas ocultas é dependente dos dados e não há uma regra geral definida (ARAS, 2017). No entanto, foi realizada a variação independente do número de neurônios de cada camada

individualmente, ou seja, para cada camada foi iniciada com um neurônio e medido o MAPE na base de dados de treinamento, em seguida foi adicionado neurônio a neurônio para avaliar a queda no MAPE, assim foi definido a quantidade de neurônios que deixam as camadas com uma quantidade de neurônios suficientes para convergência do erro na camada de saída. Os resultados estão apresentados pelo gráfico representado pela figura nove.



**Figura 9.** Evolução do MAPE com o aumento do número de neurônios por camada.

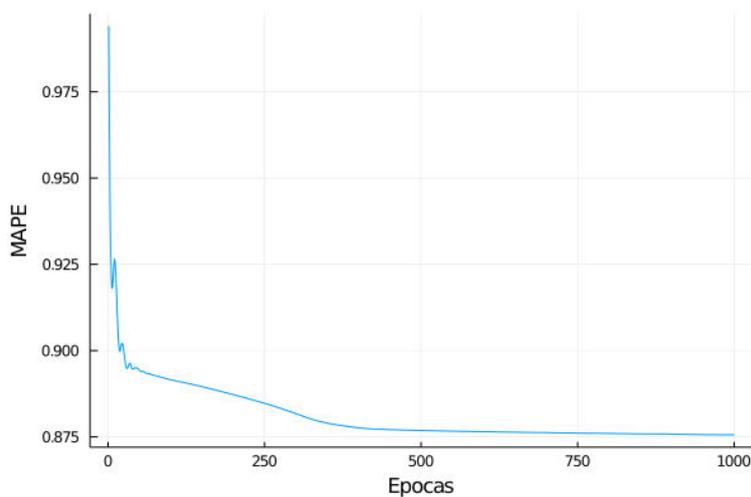
**Fonte:** Elaborado pelo autor (2021)

É possível verificar que, após um determinado ponto, o acréscimo de neurônios em cada camada não reflete em uma melhora na previsibilidade do modelo. Portanto, foi definida para o modelo a quantidade de neurônios com o melhor resultado verificado pelo gráfico da figura 9 e relacionados na tabela quatro.

**Tabela 4.** Parâmetros do modelo de Rede Neural Artificial.

	<b>Primeira Camada</b>	<b>Camada Oculta</b>	<b>Camada de Saída</b>
<b>Nº de entradas</b>	8	20	20
<b>Nº de saídas</b>	20	20	1
<b>Nº de épocas</b>	750		

Com esta configuração, foi avaliada a evolução no aprendizado pelo modelo de rede neural artificial a cada ciclo de época no gráfico da figura 10, em que o eixo *x* corresponde a quantidade de épocas e o eixo *y* a métrica de desempenho MAPE. O que evidencia que o modelo precisa de aproximadamente 750 épocas para estabilizar o erro na previsão destes dados.

**Figura 10.** Evolução do MAPE por época implementada pelo algoritmo de rede neural artificial.

**Fonte:** Elaborado pelo autor (2021)

### 4.3 COMPARATIVO DAS MÉTRICAS DE DESEMPENHO

O modelo de rede neural foi implementado inicialmente sem a clusterização e posteriormente com a clusterização. Os resultados obtidos com a implementação do modelo proposto estão indicados nas tabelas cinco, seis e sete.

Quando o modelo RNA é aplicado a cada *cluster*, os dados apresentam resultados melhores do MAPE em comparação com a aplicação do modelo sem a clusterização. Os *clusters* 1 e 3 apresentaram um resultado mais próximo aos dados não clusterizados, com valores do MAPE de 83,64% para o *cluster* 1 e 82,72% para o *cluster* 3, enquanto nos dados não clusterizados, o MAPE observado é 89,26%. Os dados dos *clusters* 1 e 3 são caracterizados pela venda *offline*.

**Tabela 5.** MAPE obtido no modelo RNA proposto.

MAPE	Dados treinamento (70%)	Dados teste (30%)
sem clusterização	87,79%	89,26%
C1	77,75%	82,72%
C2	74,54%	77,13%
C3	83,75%	83,64%
C4	69,87%	73,05%

Pelo indicador MAPE, os melhores resultados obtidos pelo modelo RNA foram nos dados dos *clusters* 2 e 4, que contemplam vendas *online* com os valores de 77,13% e 73,05% respectivamente. Entretanto, a clusterização se mostrou eficiente para tratar grupos de produtos com características de vendas semelhantes para resolver a previsão de demanda.

De forma similar, Aras *et al.* (2017), adotam um modelo de RNA, porém somente com dados históricos de venda, referentes a de dez produtos em um varejo de mobília.

Utilizando apenas este atributo interno, o autor obtém valores do MAPE entre 14,76% e 52,79%, dependendo do item, analisados de forma isolada.

Já, Huber *et al.* (2019), em seu estudo de previsão de demanda no segmento de padarias, que utiliza atributos internos e externos de forma similar a este estudo, porém somente dados de um produto, obtém um valor de MAPE para seu modelo de RNA de 27,87% e conclui no seu estudo que o RNA é um modelo que melhora a performance de previsibilidade à medida que os dados aumentam.

De forma similar, Caglayan *et al.* (2020) também adotam atributos internos e externos para sua pesquisa no segmento varejista de moda, sendo os atributos externos a temperatura e feriados. Os dados foram referentes somente a família de um modelo de sapato em uma loja de *shopping center*, sem a presença de sazonalidade, não considerando as vendas em outras filiais não localizadas em *shopping center*, por apresentarem uma maior variabilidade nos dados. O autor optou, ainda, por excluir os *outliers* antes da submissão ao seu modelo RNA. O valor do MAPE obtido foi de 8,79%.

Também com abordagem no segmento de moda, Loureiro *et al.* (2018) utilizam atributos internos referentes às características físicas dos produtos para previsão de demanda para introdução de novos produtos, com MAPE de 38,7%.

Todavia, Glaeser *et al.* (2019) utilizam informações de geolocalização, tanto para atributos internos quanto para atributos externos como, por exemplo, a localização da loja física, total da população, gênero, quantidade de residências e renda média. Estes dados são utilizados para previsão de demanda por ponto de venda através de um modelo baseado em RNA, que obtém um MAPE de 99%.

Quando comparado pelo MAE, que mede a magnitude do erro da predição, sem considerar a direção, o modelo de RNA com clusterização, apresenta uma melhora

significativa para 3 de 4 *clusters*, somente o *cluster* 1, que representa os produtos alcoólicos com venda na loja física, demonstra um valor MAE maior que os dados não clusterizados, indicando que estes produtos apresentam uma variabilidade maior na demanda *offline*, porém uma demanda mais equilibrada nas vendas *online*, que formam o *cluster* 2 para produtos alcoólicos, conforme a tabela seis.

**Tabela 6.** MAE obtido no modelo RNA proposto.

MAE	Dados treinamento (70%)	Dados teste (30%)
sem clusterização	145,9541	145,6954
C1	260,4119	243,2706
C2	34,4426	32,6827
C3	88,0675	77,9908
C4	9,4312	9,8656

No estudo de Aras *et al.* (2017), seu modelo de RNA obtém valores do MAE entre 5,20 e 261,10, para os dez itens analisados de forma isolada. Enquanto que Huber *et al.* (2019), em seu estudo no segmento de padarias, obtém um valor de MAE para seu modelo de RNA de 9,16.

Giri *et al.* (2019), abordam o segmento de varejo de moda, que utilizam no modelo de redes neurais, atributos de imagem, convertidos em vetores de informações, combinados com histórico de vendas de 31 produtos, com as transações agregadas semanalmente, para previsão de demanda de novas peças no comércio *online*. Assim, é obtido um MAE de 13,197. Já Loureiro *et al.* (2018), obtém um MAE de 1110,1 também no varejo de moda.

O indicador RMSE pontua a média da raiz da diferença elevado ao quadrado da previsão com os dados reais, o que implica em uma sensibilidade maior à medida que o erro aumenta. O modelo de RNA com clusterização, de forma similar ao indicador MAE,

também apresenta valores do RMSE melhores para três de quatro *clusters*, sendo o *cluster* 1 que apresenta o indicador RMSE maior do que os dados não clusterizados e demais *clusters*, conforme indicados na tabela sete.

**Tabela 7.** RMSE obtido no modelo RNA proposto.

RMSE	Dados treinamento (70%)	Dados teste (30%)
sem clusterização	273,7824	270,4488
C1	418,9502	335,3505
C2	61,2263	49,9936
C3	126,8745	108,7441
C4	14,7636	12,8282

No estudo de Aras *et al.* (2017), seu modelo de RNA obtém valores do RMSE entre 6,80 e 342,70, para os dez itens analisados de forma isolada. Enquanto que Huber *et al.* (2019), em seu estudo no segmento de padarias, obtém um valor de RMSE para seu modelo de RNA de 20,09.

No segmento de moda, Giri *et al.* (2019) obtém um RMSE de 19,40, enquanto que Loureiro *et al.* (2018), obtém um valor de 1909,40 para este mesmo indicador. Por fim, Boyarshinov (2019), utiliza dados de uma loja física referente a venda da categoria de macarrão para obter uma previsão de demanda através de um modelo de *machine learning* que atinge um RMSE de 1114.

Um resumo comparativo dos resultados obtidos está apresentado na tabela oito.

**Tabela 8.** Resumo comparativo das medidas de performance para previsão de demanda.

Autor	Online	Offline	MAPE	MAE	RMSE
<b>Linhares (2021)</b>	x	x	73,05% – 83,64%	9,8656 – 243,2706	12,8282 – 335,3505
<b>Aras et al. (2017)</b>		x	14,76% – 52,79%	5,20 – 241,10	6,80 – 342,70
<b>Huber et al. (2019)</b>		x	27,87%	9,16	20,09
<b>Caglayan et al. (2020)</b>		x	8,79%		
<b>Loureiro et al. (2018)</b>		x	38,70%	1110,10	1909,40
<b>Glaeser et al. (2019)</b>	x		99,00%		
<b>Giri et al. (2019)</b>	x			13,197	19,40
<b>Boyarshinov (2019)</b>		x			1114,00

Com relação à delimitação dos objetos de estudo, Aras *et al.* (2017) utilizam o histórico de vendas de dez produtos para treinamento do seu modelo no canal *offline*. Boyarshinov (2019), também utiliza apenas atributos internos de uma categoria de produtos para implementação do seu modelo no canal *offline*.

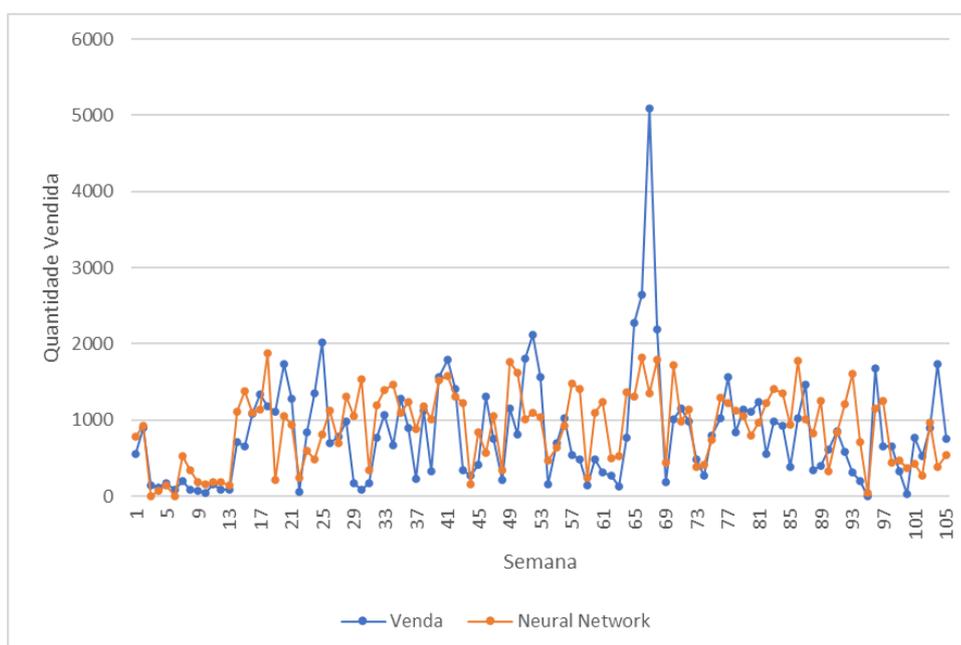
Huber *et al.* (2019) utilizam atributos internos e externos de transações *offline*, porém com dados de somente um item. Já Caglayan *et al.* (2020), também utilizam atributos internos e externos, incluindo informações de temperatura e feriados, porém se restringem aos dados da loja da rede (*offline*) que possui a menor variabilidade nos dados e apenas à uma família de produtos, conforme apontam os autores. Similarmente, Glaeser *et al.* (2019) também adotam atributos internos e externos de transações *online*, incluindo dados de georreferenciamento para previsão de vendas agregada por loja física.

Já Loureiro *et al.* (2018) utilizam características físicas dos produtos para previsão de demanda de itens novos no canal *offline*, Giri *et al.* (2019) possui uma abordagem similar, porém para introdução de itens no canal *online*.

Destes autores, somente Glaeser *et al.* (2019) e Giri *et al.* (2019) estudam o canal *online*, os demais tem o foco no canal *offline*. Portanto este estudo contribui para o estado

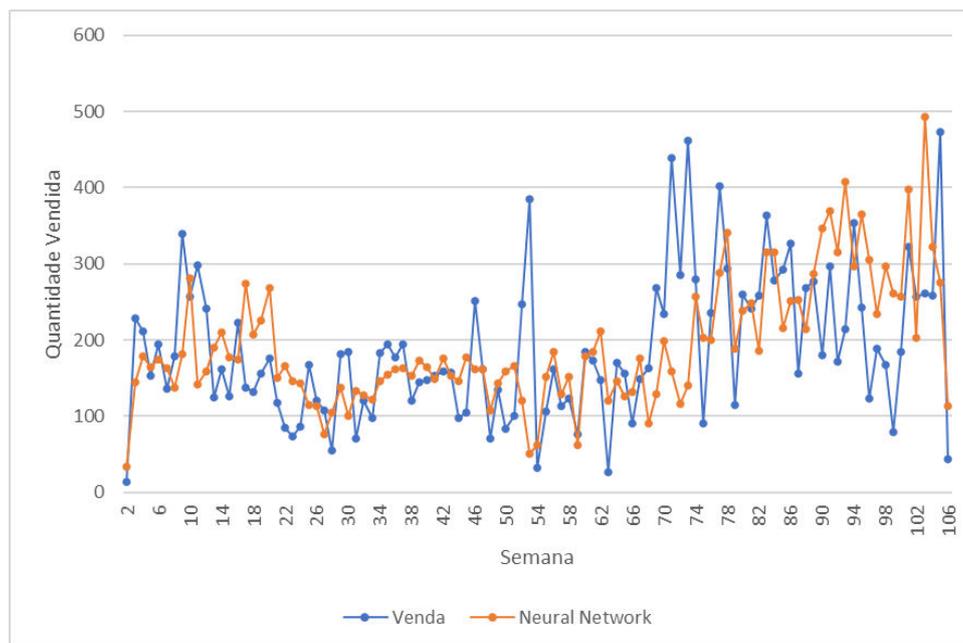
da arte ao propor e implementar um modelo com a combinação de dois algoritmos de *machine learning* em dados com atributos internos e externos, referentes à produtos de diferentes categorias (águas, sucos, chás, isotônicos, refrigerantes e cervejas) em um cenário de omnicanalidade. O modelo foi capaz de segmentar os produtos com características similares, de forma não supervisionada e com isso, aprimorar a previsão de demanda gerada pela rede neural, esta, de forma supervisionada.

Após o treinamento da rede, portanto, foi verificado que o algoritmo RNN pode ser utilizado para a previsão de demanda de produtos individuais. As figuras onze e doze, demonstram a quantidade de vendas sobrepostas pela quantidade prevista pelo modelo. A figura onze representa a evolução semanal de vendas da bebida alcoólica mais comercializada no período (cerveja) e a figura doze representa a evolução semanal das vendas do produto não alcoólico com maior venda (água tônica).



**Figura 11.** Evolução semanal de vendas da bebida alcoólica com maior venda.

**Fonte:** Elaborado pelo autor (2021)



**Figura 12.** Evolução semanal de vendas da bebida não alcoólica com maior venda.

**Fonte:** Elaborado pelo autor (2021)

A partir desta aplicação, com uma previsão de demanda que utiliza mais informações, além do histórico de vendas, é possível inferir algumas implicações gerenciais como *insights* que nesta pesquisa são tratadas na próxima seção.

#### 4.4 IMPLICAÇÕES GERENCIAIS

Como o modelo apresentado demonstrou métricas de desempenho similares à literatura, é possível utilizá-lo para obter, a partir dos dados reais de um varejo em ambiente de omnicanalidade, uma previsão de demanda dos produtos com algumas implicações gerenciais, uma vez que existe uma limitação da capacidade de cada canal, seja por restrição de espaço em prateleiras em lojas físicas ou capacidade de armazenamento em um centro de distribuição. Portanto, uma previsão de demanda assertiva por canal implica em: (i) uma base para definição de política de alocação (quantidade do produto a ser alocado em cada canal); (ii) definição de política de ressuprimento (política de reposição) que podem antecipar a demanda por canal; (iii) reduzir custos com transporte logístico; (iv) reduzir o excesso de estoque e perda de vendas por falta de estoque no canal; (v) elaborar campanhas de marketing para atingir

de forma mais assertiva cada canal de vendas; (vi) reduzir o tempo da jornada de compras e, por fim; (vii) aumentar o nível de serviço prestado ao consumidor.

## 5 CONSIDERAÇÕES FINAIS

A utilização de dados disponíveis com atributos de natureza diferente se faz relevante para previsão de demanda em um varejo com a complexidade de um ambiente de omnicanalidade. Portanto, buscou-se identificar uma oportunidade na literatura para propor um modelo baseado em inteligência artificial que possa atender à diferentes cenários para tornar a previsibilidade mais assertiva e sustentar decisões gerenciais das operações.

Este trabalho valida o modelo que combina dois algoritmos de *machine learning* que agrupa os dados que possuem atributos semelhantes (*hierarchical clustering*) seguido de uma rede neural artificial para previsão de demanda com valores dos indicadores de previsibilidade MAPE, MAE e RMSE dentro da faixa observada na revisão de literatura e, em alguns casos, com valores melhores, mesmo com uma delimitação mais ampla nos dados deste estudo.

Esta pesquisa é pioneira ao atingir o objetivo na utilização dos dados com atributos internos e externos de diversas categorias de produtos e com a presença de omnicanalidade, o que contribui para que o modelo proposto possa ser implementado em cenários semelhantes no varejo com omnicanalidade de vendas recorrentes nos canais *online* e *offline*.

Os próximos passos a serem seguidos seriam: aplicar o modelo proposto em bancos de dados ainda maiores e de segmentos diversos; considerar a otimização dos parâmetros do algoritmo de rede neural, a fim de melhorar os indicadores de previsibilidade de demanda; utilizar outros atributos relacionados ao perfil dos clientes que realizam as

transações como gênero, idade, renda média, reconhecimento facial e outros que contribuam para uma previsibilidade mais assertiva; pesquisar outros algoritmos que podem ser utilizados em um ambiente de omnicanalidade e conseqüente maior complexidade nos dados para previsão de demanda; além de calcular outros indicadores de comparação entre os modelos.

## RECONHECIMENTO

Este estudo obteve suporte financeiro da CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

## REFERÊNCIAS

Abreu, M. J. and Miranda, D. (2018) ‘The Omni-channel Concept: A Case Study about the Union of Sales Channels in Portugal’, in **IOP Conference Series: Materials Science and Engineering**.

Adamowski, J. et al. (2012) ‘Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada’, **Water Resources Research**, 48(1), pp. 1–14.

Aras, S., Deveci Kocakoç, İ. and Polat, C. (2017) ‘Comparative study on retail sales forecasting between single and combination methods’, **Journal of Business Economics and Management**, 18(5), pp. 803–832.

Aye, G. C. et al. (2015) ‘Forecasting aggregate retail sales: The case of South Africa’, **International Journal of Production Economics**. Elsevier, 160, pp. 66–79.

Baquero D.M., Mosquera-Recalde G., A.-S. S. V. (2020) ‘A New Methodology to Forecast and Manage Inventory in Mobile Warehouses’, in **Intelligent Systems Reference Library**, pp. 157–190.

Bell, D., Gallino, S. and Moreno, A. (2015) ‘Showrooms and Information Provision in Omni-channel Retail’, **Production and Operations Management**, 24(3), pp. 360–362.

Berman, B. (2019) ‘Flatlined: Combatting the death of retail stores’, **Business Horizons**. USA: ‘Kelley School of Business, Indiana University’, 62(1), pp. 75–82.

Bezanson, J. et al. (2017) ‘Julia: A Fresh Approach to Numerical Computing’, **SIAM Review**, 59(1), pp. 65–98.

Bhardwaj, G. and Swanson, N. R. (2006) 'An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series', **Journal of Econometrics**, 131(1–2), pp. 539–578.

Bielecki, A. (2019) Models of neurons and perceptrons: Selected problems and challenges, **Studies in Computational Intelligence**.

Blanco, A. P. (2018) Artificial Intelligence and the Consumer Packaged Goods Supply Chain, **Massachusetts Institute of Technology**.

Böse, J.-H. et al. (2017) 'Probabilistic demand forecasting at scale', **Proceedings of the VLDB Endowment**, 10(12), pp. 1694–1705.

Boyarshinov, V. (2019) 'Methods of machine learning for censored demand prediction', in **Springer Verlag**, p. 414.

Breiman, L. (2001) 'Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)', **Statistical Science**, 16(3), pp. 199–231.

Caglayan, N., Satoglu, S. I. and Kapukaya, E. N. (2020) 'Sales Forecasting by Artificial Neural Networks for the Apparel Retail Chain Stores', in Kahraman, C. et al. (eds) **Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making**. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing), pp. 451–456.

Cao, G. and Wu, L. (2016) 'Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting', **Energy**. Elsevier Ltd, 115, pp. 734–745.

Chawla, A. et al. (2019) 'Demand Forecasting Using Artificial Neural Networks—A Case Study of American Retail Corporation', in **Advances in Intelligent Systems and**

**Computing**. Springer Singapore, pp. 79–89.

Chen, D. et al. (2015) An Intelligent Forecasting Model for Commodities in Retail Stores, Liss 2014. Edited by Z. Zhang et al. Berlin, Heidelberg: **Springer Berlin Heidelberg**.

Chen, I.-F. and Lu, C.-J. (2017) ‘Sales forecasting by combining clustering and machine-learning techniques for computer retailing’, **Neural Computing and Applications**. Springer London, 28(9), pp. 2633–2647.

Coles, J. B. (2017) Cross-Channel Predictive Analytics for Retail Distribution Decisions. **Massachusetts Institute of Technology**.

Dutton, D. M. and Conroy, G. V. (1996) ‘A review of machine learning’, **Machine Learning**, pp. 341–367.

Efron, B. and Tibshirani, R. J. (1993) An Introduction to the Bootstrap, An introduction to Bootstrap.

Erik Brynjolfsson, Yu Jeffrey Hu, M. S. R. (2013) ‘Competing in the age of Omnichannel Retailing’, **MIT Sloan Management Review**. USA.

Fairchild, A. M. (2014) ‘Extending the Network: Defining Product Delivery Partnering Preferences for Omni-channel Commerce’, **Procedia Technology**. Elsevier B.V., 16, pp. 447–451.

Faloutsos, C. et al. (2018) ‘Forecasting big time series: Old and new’, **Proceedings of the VLDB Endowment**, 11(12), pp. 2102–2105.

Ferreira, K. J., Lee, B. H. A. and Simchi-Levi, D. (2016) ‘Analytics for an Online Retailer: Demand Forecasting and Price Optimization’, **Manufacturing & Service Operations Management**, 18(1), pp. 69–88.

Fildes, R., Ma, S. and Kolassa, S. (2019) 'Retail forecasting: Research and practice', **International Journal of Forecasting**. Elsevier B.V., (xxxx).

Giri, C. et al. (2019) 'Forecasting New Apparel Sales Using Deep Learning and Nonlinear Neural Network Regression', in 2019 **International Conference on Engineering, Science, and Industrial Applications (ICESI)**. IEEE, pp. 1–6.

Glaeser, C. K., Fisher, M. and Su, X. (2019) 'Optimal Retail Location: Empirical Methodology and Application to Practice', **Manufacturing & Service Operations Management**, 21(1), pp. 86–102.

Goh, A. T. C. (1995) 'Back-propagation neural networks for modeling complex systems', **Artificial Intelligence in Engineering**, 9(3), pp. 143–151.

Guo, Z. X., Li, M. and Wong, W. K. (2012) 'Intelligent multivariate sales forecasting using wrapper approach and neural networks', in **IEEE 10th International Conference on Industrial Informatics**. IEEE, pp. 145–150.

Hair, J. F. J. et al. (2018) *Multivariate Data Analysis*, Multivariate Data Analysis.

Hansen, Casper. Activation Functions Explained - GELU, SELU, ELU, ReLU and more. **Mlfromscratch**, 2019. Disponível em: <https://mlfromscratch.com/activation-functions-explained/#/>. Acesso em: 21 de out. de 2021.

Haykin, S. (2009) *Neural Networks and Learning Machines*. 3rd edn, **Prentice Hall**. 3rd edn. Prentice Hall.

Huang, G.-B., Zhu, Q.-Y. and Siew, C.-K. (2006) 'Extreme learning machine: Theory and applications', **Neurocomputing**, 70(1–3), pp. 489–501.

Huber, J. et al. (2019) 'A data-driven newsvendor problem: From data to decision', **European Journal of Operational Research**. Elsevier B.V., 278(3), pp. 904–915.

Hübner, A., Kuhn, H. and Wollenburg, J. (2016) 'Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework', **International Journal of Retail and Distribution Management**. Edited by H. Kotzab, Xavier Brusset and Christop, 44(3), pp. 228–247.

Hübner, A., Wollenburg, J. and Holzapfel, A. (2016) 'Retail logistics in the transition from multi-channel to omni-channel', **International Journal of Physical Distribution & Logistics Management**. Edited by C. Mena and Michael Bourlakis, 46(6/7), pp. 562–583.

Innes, M. et al. (2018) 'Fashionable Modelling with Flux'.

Islek, I. and Oguducu, S. G. (2015) 'A retail demand forecasting model based on data mining techniques', in **2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)**. IEEE, pp. 55–60.

Kharfan, M. V. W. K. C. (2018) Forecasting Seasonal Footwear Demand Using Machine Learning. **Massachusetts Institute of Technology**.

Kim, H., Lee, D. and Ryu, M. H. (2018) 'An optimal strategic business model for small businesses using online platforms', **Sustainability** (Switzerland), 10(3), pp. 1–11.

Kingma, D. P. and Ba, J. L. (2015) 'Adam: A method for stochastic optimization', in **3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings**, pp. 1–15.

Kumar, V. et al. (2019) 'Customer engagement in service', **Journal of the Academy of Marketing Science**. Journal of the Academy of Marketing Science, 47(1), pp. 138–160.

Lagorio, A., Pinto, R. and Golini, R. (2016) 'Research in urban logistics: a systematic literature review', **International Journal of Physical Distribution & Logistics**

**Management.** Wiesbaden: VS Verlag für Sozialwissenschaften, 46(10), pp. 908–931.

Lau, K. H. (2012) ‘Demand management in downstream wholesale and retail distribution: A case study’, **Supply Chain Management**, 17(6), pp. 638–654.

Lee, S. H. and Jung, S. (2020) ‘Fashion consumers’ channel-hopping profiles by psychographics and demographics’, **International Journal of Market Research**, 62(5), pp. 615–632.

Livieris, I. E. et al. (2019) ‘Employing Constrained Neural Networks for Forecasting New Product’s Sales Increase’, in **IFIP Advances in Information and Communication Technology**, pp. 161–172.

Loureiro, A. L. D., Miguéis, V. L. and da Silva, L. F. M. (2018) ‘Exploring the use of deep neural networks for sales forecasting in fashion retail’, **Decision Support Systems**. Elsevier B.V, 114, pp. 81–93.

Marchet, G. et al. (2018) ‘Business logistics models in omni-channel: a classification framework and empirical analysis’, **International Journal of Physical Distribution & Logistics Management**, 48(4), pp. 439–464.

Martino, G. et al. (2019) **Business Models and ICT Technologies for the Fashion Supply Chain**. Edited by R. Rinaldi and R. Bandinelli. Cham: Springer International Publishing (Lecture Notes in Electrical Engineering).

Matias B., A. (2018) Improving Complex Sale Cycles and Performance by using Machine Learning and Predictive Analytics to Understand the Customer Journey. **Massachusetts Institute of Technology**.

Melacini, M. et al. (2018) ‘E-fulfilment and distribution in omni-channel retailing: a systematic literature review’, **International Journal of Physical Distribution &**

**Logistics Management**, 48(4), pp. 391–414.

Paul, J. et al. (2019) ‘Shared Capacity Routing Problem – An Omni-channel Retail Study’, **European Journal of Operational Research**, 273(2), pp. 731–739.

Di Pillo, G. et al. (2016) ‘An application of support vector machines to sales forecasting under promotions’, **4OR. Springer Berlin Heidelberg**, 14(3), pp. 309–325.

Piotrowicz, W. and Cuthbertson, R. (2014) ‘Introduction to the special issue information technology in retail: Toward omnichannel retailing’, **International Journal of Electronic Commerce**, pp. 5–16.

Punj, G. and Stewart, D. W. (1983) ‘Cluster Analysis in Marketing Research: Review and Suggestions for Application’, **Journal of Marketing Research**, 20(2), p. 134.

Qu, T. et al. (2017) ‘Demand prediction and price optimization for semi-luxury supermarket segment’, **Computers & Industrial Engineering**, 113, pp. 91–102.

Qureshi, F. F. et al. (2017) ‘Integration of OMNI channels and machine learning with smart technologies’, **Journal of Ambient Intelligence and Humanized Computing**. Springer Berlin Heidelberg, 0(0), p. 0.

Rigby, D. K. (2011) ‘The Future of Shopping’, **Harvard Business Review**.

Rosenblatt, F. (1957) ‘The Perceptron - A Perceiving and Recognizing Automaton’, **Report 85, Cornell Aeronautical Laboratory**, pp. 460–1.

Rumelhart, D. E., Feldman, J. A. and Hayes, P. J. (1986) **Parallel Distributed Processing computational model of cognition and perception, Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 1: Foundations**.

Saghiri, S. et al. (2017) ‘Toward a three-dimensional framework for omni-channel’,

**Journal of Business Research**, 77, pp. 53–67.

Salami, M., Sobhani, F. M. and Ghazizadeh, M. S. (2018) ‘Short-term forecasting of electricity supply and demand by using the wavelet-PSO-NNS-SO technique for searching in big data of iran’s electricity market’, **Data**, 3(4).

Salinas, D. et al. (2019) ‘DeepAR: Probabilistic forecasting with autoregressive recurrent networks’, **International Journal of Forecasting**. Elsevier B.V., (xxxx).

Semenov, V. P., Chernokulsky, V. V. and Razmochaeva, N. V. (2017) ‘Research of artificial intelligence in the retail management problems’, in **2017 IEEE II International Conference on Control in Technical Systems (CTS)**. IEEE, pp. 333–336.

Steinhoff, L. et al. (2019) ‘Online relationship marketing’, **Journal of the Academy of Marketing Science**. Journal of the Academy of Marketing Science, 47(3), pp. 369–393.

Tee, J. E. C. et al. (2017) ‘Day-ahead forecasting of wholesale electricity pricing using extreme learning machine’, in **TENCON 2017 - 2017 IEEE Region 10 Conference**. IEEE, pp. 2973–2977.

Thobani, S. (2018) **Improving E-Commerce Sales Using Machine Learning**, Massachusetts Institute of Technology.

da Veiga, C. P. et al. (2016) ‘Demand forecasting based on natural computing approaches applied to the foodstuff retail segment’, **Journal of Retailing and Consumer Services**. Elsevier, 31, pp. 174–181.

Verhoef, Peter C., Kannan, P. K. and Inman, J. J. (2015) ‘From Multi-Channel Retailing to Omni-Channel Retailing’, **Journal of Retailing**. New York University, 91(2), pp. 174–181.

Verhoef, Peter C, Kannan, P. K. and Inman, J. J. (2015) ‘From Multi-Channel Retailing

to Omni-Channel Retailing Introduction to the Special Issue on Multi-Channel Retailing &’, **Journal of Retailing**. New York University, 91(2), pp. 174–181.

Verstraete, G., Aghezzaf, E. H. and Desmet, B. (2019) ‘A data-driven framework for predicting weather impact on high-volume low-margin retail products’, **Journal of Retailing and Consumer Services**. Elsevier Ltd, 48(December 2018), pp. 169–177.

Wang, L., Fan, H. and Wang, Y. (2018) ‘Sustainability Analysis and Market Demand Estimation in the Retail Industry through a Convolutional Neural Network’, **Sustainability**, 10(6), p. 1762.

Wong, W. K. and Guo, Z. X. (2010) ‘A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm’, **International Journal of Production Economics**. Elsevier, 128(2), pp. 614–624.

Wong, W. K. and Guo, Z. X. (2013) ‘Intelligent sales forecasting for fashion retailing using harmony search algorithms and extreme learning machines’, in **Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI)**. Elsevier, pp. 170–195.

Xue, H. et al. (2018) ‘Research on demand forecasting of retail supply chain emergency logistics based on NRS-GA-SVM’, in **2018 Chinese Control And Decision Conference (CCDC)**. IEEE, pp. 3647–3652.

Yue, L. et al. (2010) ‘Selective and Heterogeneous SVM Ensemble for Demand Forecasting’, in **2010 10th IEEE International Conference on Computer and Information Technology**. IEEE, pp. 1519–1524.

Zhang, G., Eddy Patuwo, B. and Y. Hu, M. (1998) ‘Forecasting with artificial neural

networks', **International Journal of Forecasting**, 14(1), pp. 35–62.