

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ - UNIOESTE
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS, GESTÃO E
SUSTENTABILIDADE – PPGTGS (MESTRADO PROFISSIONAL)

RODRIGO PEREIRA FONTES

**MODELO PREDITIVO DE VENDAS PARA UMA FRANQUIA
DE *BUBBLE TEA* PARA RECOMENDAÇÃO DE ESTOQUE**

FOZ DO IGUAÇU

2020

RODRIGO PEREIRA FONTES

**MODELO PREDITIVO DE VENDAS PARA UMA
FRANQUIA DE *BUBBLE TEA* PARA RECOMENDAÇÃO DE
ESTOQUE**

Dissertação apresentada ao **Programa de Pós-Graduação em Tecnologias, Gestão e Sustentabilidade** da Universidade Estadual do Oeste do Paraná, como requisito parcial para a obtenção do título de **Mestre**. Área de Concentração: Tecnologia e Gestão.

Orientador: Prof. Dr. Clodis Boscaroli

FOZ DO IGUAÇU

2020

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Fontes, Rodrigo Pereira

Modelo preditivo de vendas para uma franquia de bubble tea para recomendação de estoque / Rodrigo Pereira Fontes; orientador(a), Clodis Boscarioli , 2020.

120 f.

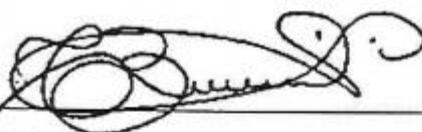
Dissertação (mestrado profissional), Universidade Estadual do Oeste do Paraná, Campus de Foz do IguaçuCentro de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Tecnologias, Gestão e Sustentabilidade, 2020.

1. Gestão de estoques. 2. Previsão de demandas. 3. Aprendizagem de máquina. 4. Séries temporais. I. , Clodis Boscarioli. II. Título.

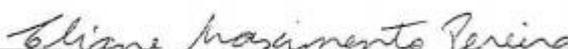
RODRIGO PEREIRA FONTES

**MODELO PREDITIVO DE VENDAS PARA UMA FRANQUIA DE
BUBBLE TEA PARA RECOMENDAÇÃO DE ESTOQUE**

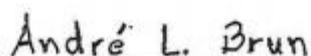
Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Tecnologias, Gestão e Sustentabilidade - PGTGS da Universidade Estadual do Oeste do Paraná, aprovado pela Banca Examinadora:



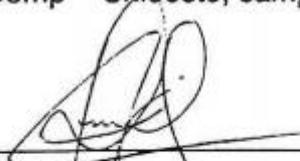
Prof. Dr. Clodis Boscarioli - Orientador
Professor do PGTGS – Unioeste, campus de Foz do Iguaçu



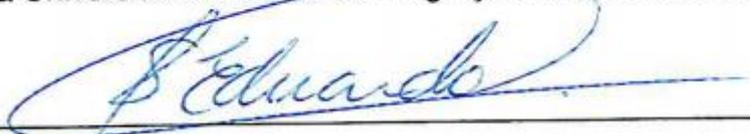
Prof.ª Dr.ª. Eliane Nascimento Pereira
Professora do PGTGS – Unioeste, campus de Foz do Iguaçu



Prof. Dr. André L. Brun
Professor do PPGComp – Unioeste, campus de Cascavel



Prof. Dr. Marcelo Nepomoceno Kapp
Professor da Universidade Federal da Integração Latino Americana



Prof. Dr. Eduardo Cesar Dechechi
Coord. do Mestrado Profissional em Tecnologias, Gestão e Sustentabilidade

Portaria Nº3279/2018-GRE - UNIOESTE – Campus de Foz do Iguaçu

Foz do Iguaçu, 30 de julho de 2020

Aos meus pais, Vicente e Terezinha, por me apoiarem de forma incondicional desde sempre, dedicando seus esforços integralmente para os filhos.

À minha amada esposa Andressa Rospirski, por seu apoio e contribuições fundamentais para que fosse possível chegar até aqui.

Aos meus irmãos, por me incentivarem a buscar bens de valor inestimável: cultura e sabedoria.

AGRADECIMENTOS

Ao professor Clodis Boscarioli, pela orientação exigindo nada menos que o meu melhor. Foram muitas horas de conversas, compartilhamento e aprendizagem. Um professor que além de orientador, durante todo o processo, se mostrou um amigo com diversos conselhos valiosos.

Aos professores do PPGTGS, que me proporcionaram a oportunidade de aprender e acumular novos conhecimentos dos quais eu nem sonhava que existiam.

Aos membros da Banca Examinadora que doaram seu precioso tempo para avaliar e contribuir com experiência e conhecimento para tornar essa dissertação ainda mais rica.

Aos colegas do PPGTS, que compartilhavam das dificuldades e das vitórias, sempre estendendo a mão para ajudar sempre que precisei.

A Bubble Mix, sob a figura de seu Diretor Administrativo, Rodrigo Balotin, que abriu as portas desde o início, além de incentivar e acompanhar tudo o que foi realizado nessa pesquisa.

Aos meus amigos e familiares, que tiveram que suportar muitas ausências e períodos de distanciamento enquanto eu dedicava o tempo que me sobrava para que fosse possível chegar a este ponto. Muito obrigado pelo apoio e carinho, foram determinantes para que esse grande objetivo fosse alcançado.

Ao meu pai e minha mãe, por palavras de incentivo, elogios e sempre me fazendo lembrar do porquê de se fazer as coisas. Me permitiram manter o legado e a herança mais valiosa da nossa família: a perseverança.

A minha esposa Andressa Rospirski, que sempre esteve presente para suprir as minhas ausências em nossos compromissos, por entender a importância que esta experiência teve na minha vida. Obrigado pelo seu amor e dedicação, sem eles eu provavelmente não teria conseguido.

Encerro esses agradecimentos deixando o meu muito obrigado para todos, que direta ou indiretamente, participaram desta pesquisa.

“Não é necessário que você veja toda a escada. Apenas dê o primeiro passo”.

Martin Luther King

RESUMO

FONTES, R. P. (2020). *Modelo preditivo de vendas para uma franquia de bubble tea para recomendação de estoque*. Dissertação de Mestrado – Programa de Pós-Graduação em Tecnologias, Gestão e Sustentabilidade - PGTGS, Universidade Estadual do Oeste do Paraná – UNIOESTE, Foz do Iguaçu, Paraná, Brasil.

As organizações estão em constante busca pela sustentabilidade por meio da oferta e demanda. Os estoques representam componentes significativos e devem ser gerenciados de forma que os recursos financeiros sejam empregados da maneira mais racional possível. Uma gestão ineficiente de estoques causa efeitos indesejados como a imobilização de capital e o aumento de custos de manutenção de estoques. Mesmo com o avanço das tecnologias computacionais, nem todas as empresas brasileiras utilizam métodos quantitativos formais para a gestão do estoque, principalmente as de pequeno e médio portes, muito embora, algumas empresas utilizam para lidar com as rotinas administrativas, podem utilizar os dados relativos ao consumo de produtos para prever demandas e provisionar seu estoque. A pesquisa teve como objetivo analisar padrão de consumo de uma franquia de *bubble tea* e aplicou técnicas de predição comparando os resultados gerados. A técnica baseada em Modelos Lineares Generalizados foi a que apresentou o melhor resultado, contribuindo para o provisionamento mais adequado de estoque para atendimento da demanda, de acordo com o padrão de consumo.

Palavras-chave: Gestão de estoques, Previsão de demandas, Aprendizagem de máquina, Séries temporais.

ABSTRACT

FONTES, R. P. (2020). *Predictive sales model for a bubble tea franchise for stock recommendation*. Master's Dissertation - Postgraduate Program in Technologies, Management and Sustainability - PGTGS, State University of Western Paraná - UNIOESTE, Foz do Iguaçu, Paraná, Brazil.

Organizations are constantly seeking to ensure sustainability through supply and demand. Inventories represent significant components and must be managed in such a way that financial resources are used in the most rational way possible. Inefficient inventory management causes unwanted effects like capital immobilization and increased inventory maintenance costs. Even with the advancement of computational technologies, not all Brazilian companies use formal quantitative methods for inventory management, especially small and medium-sized ones, although some companies use it to deal with administrative routines, they can use data related to consumption of products to forecast demands and provision your inventory. The research aimed to analyze the consumption pattern of a bubble tea franchise and applied prediction techniques comparing the results generated. The machine learning technique based on Generalized Linear Models was the one that presented the best result, contributing to the most adequate provisioning of stock to meet demand, according to the consumption pattern.

Keywords: Inventory management, Forecast demand, Machine learning, Time series.

SUMÁRIO

| | | |
|--------|---|----|
| 1. | INTRODUÇÃO | 12 |
| 2. | GESTÃO E CONTROLE DE ESTOQUES..... | 17 |
| 2.1. | ANÁLISE DOS ESTOQUES..... | 18 |
| 2.2. | MODELO DE REPOSIÇÃO CONTÍNUA | 20 |
| 2.3. | MODELO DE REPOSIÇÃO PERIÓDICA | 22 |
| 2.4. | LOTE ECONÔMICO DE COMPRA (LEC)..... | 23 |
| 3. | TÉCNICAS DE PREDIÇÃO DE DEMANDAS..... | 26 |
| 3.1. | TÉCNICAS QUALITATIVAS..... | 33 |
| 3.2. | TÉCNICAS QUANTITATIVAS..... | 36 |
| 3.2.1. | Séries Temporais..... | 36 |
| 3.2.2. | Média Móvel..... | 39 |
| 3.2.3. | Suavização Exponencial | 41 |
| 3.2.4. | ARIMA..... | 42 |
| 3.3. | TÉCNICAS DE APRENDIZAGEM DE MÁQUINA | 46 |
| 3.3.1. | Redes Neurais Artificiais | 47 |
| 3.3.2. | Modelos Lineares Generalizados | 50 |
| 3.3.3. | Florestas Aleatórias..... | 51 |
| 3.4. | MÉTRICAS PARA MEDIR ACURÁCIA DE PREDIÇÕES | 54 |
| 3.4.1. | Erro Médio Absoluto | 54 |
| 3.4.2. | Raiz Quadrada do Erro Médio..... | 55 |
| 3.4.3. | Erro Percentual Absoluto Médio..... | 55 |
| 3.4.4. | Coefficiente de Determinação | 55 |
| 4. | PROCEDIMENTOS METODOLÓGICOS | 57 |
| 4.1. | ANÁLISE EXPERIMENTAL | 61 |
| 4.2. | APLICAÇÃO DAS TÉCNICAS QUANTITATIVAS SELECIONADAS | 74 |
| 4.2.1. | Aplicação da técnica de Médias Móveis..... | 75 |
| 4.2.2. | Aplicação da técnica de Suavização Exponencial | 78 |
| 4.2.3. | Aplicação do modelo ARIMA..... | 80 |
| 4.2.4. | Aplicação das Técnicas de Aprendizagem de Máquina | 84 |
| 4.2.5. | Aplicação de Redes Neurais Artificiais | 85 |
| 4.2.6. | Aplicação de Modelos Lineares Generalizados | 88 |

| | |
|---|-----|
| 4.2.7. Aplicação de Florestas Aleatórias | 93 |
| 4.3. COMPARATIVO ENTRE TÉCNICAS DE PREDIÇÃO APLICADAS | 98 |
| 5. O MODELO PREDITIVO DE ESTOQUE PROPOSTO | 101 |
| 6. CONSIDERAÇÕES FINAIS..... | 106 |
| REFERÊNCIAS..... | 109 |
| ANEXO..... | 115 |

LISTA DE SIGLAS E ABREVIATURAS

| | |
|-------|--|
| ERP | <i>Enterprise Resource Planning</i> |
| PP | Ponto de Pedido |
| LEC | Lote Econômico de Compra |
| MLP | <i>Multilayer Perceptron</i> |
| GNU | <i>General Public License</i> |
| INMET | Instituto Nacional de Meteorologia |
| MAPE | <i>Mean Absolute Percentage Error</i> |
| MAE | <i>Mean Absolute Error</i> |
| RMSE | <i>Root Mean Square Error</i> |
| EAM | Erro Absoluto Médio |
| EPAM | Erro Percentual Absoluto Médio |
| REQM | Raiz do Erro Quadrático Médio |
| ARIMA | <i>Auto Regressive Integrated Moving Average</i> |
| ARMA | <i>Auto Regressive Moving Average</i> |
| PDV | Ponto de Venda |
| XLS | <i>Excel Spreadsheet</i> |
| FAC | Função de Auto Correlação |
| FACP | Função de Auto Correlação Parcial |
| AIC | <i>Akaike Information Criteria</i> |
| RNA | Redes Neurais Artificiais |
| NNAR | <i>Neural Network Auto Regressive</i> |
| MLG | Modelos Lineares Generalizados |

GLM *Generalized Linear Models*

RF *Random Forest*

CSV *Comma Separated Values*

FTP *Ficha Técnica de Preparo*

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Movimentação de insumo no sistema de gestão | 19 |
| Figura 2 - Monitoramento dos níveis de estoque para identificação do PP | 21 |
| Figura 3 – Observação dos lotes de ressuprimento variáveis | 22 |
| Figura 4 - Processo de utilização de previsões | 27 |
| Figura 5 - <i>Framework</i> para previsão e planejamento | 31 |
| Figura 6 - Categorias de técnicas de previsão | 32 |
| Figura 7 - Tipos de séries temporais | 37 |
| Figura 8 - Processo Box & Jenkins | 42 |
| Figura 9 - Redes Neurais <i>feedforward</i> | 48 |
| Figura 10 - Arquitetura das Florestas Aleatórias..... | 53 |
| Figura 11 - Processos de desenvolvimento da pesquisa..... | 57 |
| Figura 12 - Relatório de movimentação de produtos por dia | 59 |
| Figura 13 – Processo para previsão do estoque | 60 |
| Figura 14 - Vendas de <i>bubble tea</i> por categoria..... | 62 |
| Figura 15 - Vendas mensais por categoria | 63 |
| Figura 16 - Total diário de vendas por categoria | 63 |
| Figura 17 - Dados históricos sobre o clima de Foz do Iguaçu – PR | 64 |
| Figura 18 - Dispersão da temperatura no tempo | 65 |
| Figura 19 - Dispersão das vendas, temperatura e umidade no tempo | 66 |
| Figura 20 - Vendas de acordo com a temperatura | 66 |
| Figura 21 – Recorte do conjunto de dados contendo as estações | 68 |
| Figura 22 - Vendas de bebidas por estação do ano | 69 |

| | |
|--|-----|
| Figura 23 - Estrutura com feriados e fins de semana | 70 |
| Figura 24 - Vendas em dias úteis e fins de semana | 71 |
| Figura 25 - Vendas em dias úteis e feriados..... | 72 |
| Figura 26 - Estrutura com informações sobre as férias escolares | 73 |
| Figura 27 - Vendas de bebidas durante as férias escolares | 73 |
| Figura 28 - Processo para aplicação das técnicas quantitativas | 74 |
| Figura 29 - Médias Móveis para fevereiro/2020..... | 77 |
| Figura 30 - Previsão de vendas com Suavização Exponencial | 79 |
| Figura 31 - Séries temporais vendas de <i>bubble tea</i> | 80 |
| Figura 32 - Resultado do teste de Dickey-Fuller..... | 81 |
| Figura 33 – FAC e FACP | 81 |
| Figura 34 - Previsão de vendas para fevereiro/2020 com ARIMA (0,0,0)..... | 82 |
| Figura 35 - Processo de Aprendizagem de Máquina | 85 |
| Figura 36 - Previsão de vendas diárias de <i>bubble tea</i> utilizando RNA | 86 |
| Figura 37 - Recorte do conjunto de dados gerado pelo MLG | 90 |
| Figura 38 - Total de vendas por categoria para fevereiro/2020 | 91 |
| Figura 39 – Correlação entre previsto (MLG) e realizado | 92 |
| Figura 40 - Previsão de vendas para fevereiro/2020 | 96 |
| Figura 41 - Correlação entre previsto (RF) e realizado..... | 97 |
| Figura 42 - Recorte da FTP da bebida Yakult Cranberry..... | 101 |
| Figura 43 - Recorte do arquivo CSV contendo parte da FTP dos produtos.. | 102 |
| Figura 44 - Recorte de produtos previstos sumarizados por categoria..... | 102 |
| Figura 45 - Cálculo de ingredientes para atender previsão | 103 |
| Figura 46 - Conversões dos insumos | 104 |

Figura 47 - Recorte da recomendação de estoque.....105

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 1 - Faturamento do segundo trimestre de uma empresa | 40 |
| Tabela 2 – Previsão para o terceiro trimestre de uma empresa | 40 |
| Tabela 3 - Estrutura dos dados de estações climáticas..... | 68 |
| Tabela 4 - Cálculo de Médias Móveis para vendas de fevereiro de 2020..... | 76 |
| Tabela 5 - Métricas para erros de previsão de Médias Móveis..... | 77 |
| Tabela 6 - Cálculo de Médias Móveis para vendas de fevereiro de 2020..... | 78 |
| Tabela 7 - Métricas para erros de previsão da Suavização Exponencial | 80 |
| Tabela 8 – Critério para seleção do modelo | 82 |
| Tabela 9 – Previsão ARIMA (0, 0, 0) para vendas de fevereiro de 2020 | 83 |
| Tabela 10 - Métricas para erros de previsão ARIMA (0, 0, 0)..... | 83 |
| Tabela 11 – Previsão NNAR(9, 5) para vendas de fevereiro de 2020 | 87 |
| Tabela 12 - Métricas para erros de previsão NNAR (9, 5)..... | 87 |
| Tabela 13 – Previsão MLG para vendas de fevereiro de 2020..... | 89 |
| Tabela 14 - Métricas para erros de previsão utilizando MLG..... | 89 |
| Tabela 15 – Previsão MLG por Categoria..... | 93 |
| Tabela 16 - Previsão para vendas de fevereiro de 2020 | 94 |
| Tabela 17 - Métricas para erros de previsão utilizando Florestas Aleatórias.. | 94 |
| Tabela 18 – Previsão de Florestas Aleatórias por Categoria de Bebidas | 98 |
| Tabela 19 – Comparativo entre técnicas de predição..... | 99 |
| Tabela 20 – Recursos computacionais utilizados | 100 |

1. INTRODUÇÃO

As organizações estão em constante busca da sustentabilidade econômica por meio do equilíbrio entre oferta e demanda. Para atingir eficiência e encontrar este equilíbrio, as empresas necessitam disponibilizar o produto de acordo com as necessidades do cliente, na hora certa e na quantidade correta, com os menores custos possíveis (BALLOU, 2009).

Para Viana (2010, p. 108), “Estoques representam componentes extremamente significativos, seja sob aspectos econômico-financeiros ou operacionais críticos”. O autor destaca também a importância de uma gestão efetiva de estoques ao descrever que “os materiais concorrem, quase sempre, com mais de 50% do custo do produto vendido, o que faz com que os recursos financeiros alocados a estoques devam ser empregados sob a forma mais racional possível” (VIANA, 2010, p. 108). Estas afirmações são ainda atuais e, dessa forma, torna-se cada vez mais necessária a adoção da estratégia da Excelência Operacional para tornar os processos de gestão de estoque mais eficientes e eficazes dentro da organização.

De acordo com Hayes & Upton (1998), a estratégia da Excelência Operacional é adotada por empresas que atuam em mercados nos quais a relação de qualidade frente ao preço está sob constante avaliação por parte do consumidor, além de influenciar diretamente na rentabilidade do negócio. Nesses casos, para obter uma maior probabilidade de se destacar atuando nesses mercados, geralmente adota-se uma vantagem competitiva baseada nas operações. A eficácia nas operações de rotina não serve apenas para reforçar a capacidade competitiva, pois quando incorporada em recursos humanos e nos processos operacionais da empresa, são inerentemente difíceis de serem copiadas por concorrentes. Ao otimizar os processos de gestão de estoques, é possível diminuir seu custo e aumentar a rentabilidade do negócio.

Richards (2014) define o papel do profissional responsável pela gestão do estoque afirmando que é de sua responsabilidade maximizar a eficiência de uso dos recursos operacionais enquanto satisfaz aos requisitos do cliente. O autor também cita seis princípios básicos da gestão de estoques: precisão, controle de custos, limpeza, eficiência, proteção e segurança.

No caso da empresa objeto deste estudo, que enfrenta a realidade de grande parte das empresas brasileiras, o maior desafio é manter o nível de estoque adequado para atendimento da demanda sem que haja demasia de produtos armazenados que venham a gerar custos adicionais ou imobilizar o capital de giro, mas que também não venham a faltar, uma vez que esse cenário causaria a perda de vendas e prejudicaria a reputação do negócio gerando um impacto negativo nos resultados operacionais. Como também não há uma previsão adequada para demanda dos produtos a serem comercializados, um outro desafio seria descobrir qual o nível de estoque adequado sem que seja gerado o "efeito chicote", descrito por Coelho, Follmann & Rodriguez (2008, p. 21) como:

"o resultado da discrepância entre a demanda real e a prevista unida à intenção das empresas alinharem sua oferta a essa demanda, sem deixar de atendê-la. Dessa forma, as empresas, por não possuírem a informação correta de seus clientes, buscam se proteger e garantir o nível de atendimento por meio do aumento do nível de estoques para uma possível variação nessa demanda".

Como principal consequência desse efeito, temos a imobilização de capital e o aumento de custos de manutenção do estoque. Para Giansesi & Biazzi (2011), as principais técnicas formais para provisionamento de estoques foram desenvolvidas há algumas décadas e desde então, empregou-se um esforço acadêmico na aplicação de métodos heurísticos às decisões para descobrir o que comprar, quanto comprar e quando suprir os estoques. Segundo esses autores, mesmo com o suporte oferecido por todo esse desenvolvimento, é raro encontrar empresas brasileiras que utilizem métodos quantitativos formais para a gestão do estoque, e destacam também que, ao contrário do esperado, os empresários e administradores brasileiros ainda recorrem ao uso de métodos empíricos, qualitativos e baseados em intuição (GIANESI & BIAZZI, 2011), o que ainda se aplica à realidade empresarial atual.

A empresa que servirá de base para este estudo, trata-se de uma franquia com mais de vinte lojas espalhadas no Brasil. O principal produto vendido é o *bubble tea*, uma bebida de origem taiwanesa que combina ingredientes como chás, cafés, essências de frutas, leite fermentado, pérolas de tapioca, gelatinas e outros insumos. A unidade da franquia que servirá de piloto para a pesquisa, está situada em um *shopping center* na região central da cidade de Foz do Iguaçu, Paraná.

A franqueadora recomenda ao franqueado, possuir em seu quadro ao menos quatro colaboradores que trabalham em turnos, sendo um destes o Franqueado Operador, que pode ser definido como o responsável pelo gerenciamento dos processos de trabalho, das operações, do controle de estoque e da gestão dos recursos humanos da empresa. Também é o responsável por manter os padrões estabelecidos para a rede e de responder formalmente sobre questionamentos e demais assuntos pertinentes junto à franqueadora.

Apesar de utilizar um sistema de informação para suporte aos processos, a gestão do estoque dos franqueados, de modo geral, não é eficiente, pois é totalmente dependente das informações que são adicionadas ao sistema pelo Operador da unidade, de forma inteiramente manual. O software utilizado oferece uma forma básica de controle, em que os itens comprados são adicionados manualmente à conta estoque e para cada produto vendido, os insumos são debitados. Porém, nenhuma inferência é gerada a partir deste controle, sendo responsabilidade do Operador da unidade, realizar contagens frequentes para observar os itens que precisam de reposição. O método de reposição, atualmente, é feito de maneira empírica pelo Operador, baseando-se em sua experiência na gestão da unidade, bem como em eventos derivados de sazonalidades, como promoções, feriados, férias escolares, entre outros.

Por conta desta forma de controle, todos os franqueados da rede Bubble Mix enfrentam as mesmas dificuldades na hora de realizar a compra dos insumos necessários para produção das bebidas. A franqueadora, utiliza uma informação padronizada para recomendação de compra, independente da praça onde a unidade está instalada, padrão de consumo local e outras variáveis pertinentes ao contexto de cada unidade. Ao implementar uma unidade, a franquia envia uma “lista de compras”, com os itens que devem ser adquiridos. Em algumas praças, esse estoque inicial se manterá em bom nível por alguns dias. Em outras praças, pode durar semanas. Essa situação faz com que a compra dos insumos para o estoque seja realizada de maneira empírica pelo franqueado operador de cada unidade, abrindo margem para erros como: falta de produtos, vencimento de insumos e capital imobilizado, uma vez que não existe um nível de estoque ideal pré-determinado que leve em consideração o padrão de consumo.

Na pesquisa de Sickle & Beilock (1981), os autores concluíram que a acurácia do modelo preditivo na indústria alimentícia é extremamente dependente do horizonte temporal a ser projetado. As pesquisas de Adebajo & Mann (2000) e Arunraj & Ahrens (2015), reforçam a importância da utilização de modelos preditivos nesta indústria, destacando a lucratividade das operações e a melhoria na gestão de estoques.

Armstrong (2002) cita a importância de ter previsões de curto prazo para evitar estoques redundantes ou insuficientes. Segundo Ross (2015), a previsão diária de demandas possui vantagens, uma vez que a agregação de dados históricos diários pode promover resultados superiores por conta da massa de dados disponível para criação do modelo preditivo.

Para Buzby *et al.* (2009) e Eriksson (2012), a melhoria de modelos preditivos como suporte para gestão de pedidos, contribui para a diminuição de desperdício de insumos. De acordo com Arunraj & Ahrens (2015), o padrão de consumo no ramo alimentício é específico ao cliente e ao produto. Esses autores afirmam ainda que a variação na demanda pode ser categorizada em flutuações de curto prazo, como por exemplo, feriados e promoções, padrões sazonais médios, como por exemplo, férias escolares e estação climática e tendências de longo prazo, como por exemplo, situações econômicas.

Em sintonia com os pontos observados por estes autores, o presente estudo busca propor um modelo preditivo de vendas para uma franquia de *bubble tea* utilizando dados históricos, para que posteriormente, um nível de estoque para atendimento da demanda seja recomendado, de forma que maximize a lucratividade do negócio.

Apesar da implantação em massa dos sistemas de gestão – chamados *Enterprise Resource Planning* (ERP) - muitas empresas encontram obstáculos que dificultam a alteração de seus processos de gestão de modo a utilizar o potencial dos dados armazenados por meio desses sistemas, tal que se identifica aqui a seguinte questão de pesquisa: Como a utilização dos dados relativos ao consumo de produtos pode contribuir na previsão de vendas a fim de recomendar o estoque?

Para responder esta questão, o objetivo geral da pesquisa é propor um modelo preditivo de vendas para a rede Bubble Mix, utilizando uma unidade como piloto para

a análise experimental considerando seu contexto e, por meio das informações disponíveis, recomendar um nível de estoque adequado de acordo com o padrão de consumo. A partir deste contexto, temos os seguintes objetivos específicos:

- Identificar técnicas de predição que podem ser utilizadas;
- Consolidar um conjunto de dados coletados do ERP da unidade piloto;
- Identificar o melhor modo preditivo a ser utilizado de acordo com o conjunto de dados criado;
- Recomendar o estoque baseado na previsão do melhor modelo.

Este documento segue assim organizado: O Capítulo 2 traz uma visão geral sobre gestão e controle de estoques. O Capítulo 3 introduz técnicas para predição de demandas e modelos preditivos geralmente utilizados por organizações. O Capítulo 4 apresenta a empresa, os procedimentos metodológicos adotados, a análise experimental realizada, a aplicação das técnicas de predição selecionadas e compara os resultados para definir o melhor modelo. O Capítulo 5 apresenta a proposta do modelo preditivo para recomendação de estoque. E por fim, o Capítulo 6 traz as considerações finais e perspectivas da pesquisa realizada.

2. GESTÃO E CONTROLE DE ESTOQUES

Uma das maiores preocupações da administração de materiais é controlar os níveis de estoques da organização (POZO, 2002).

Para Viana (2010, p. 117) a gestão de estoques pode ser definida como:

“conjunto de atividades que visa, por meio das respectivas políticas de estoque, ao pleno atendimento das necessidades da empresa, com a máxima eficiência e ao menor custo, através do maior giro possível para o capital investido em materiais”.

Uma gestão eficiente do estoque de uma empresa, contribui significativamente para a coordenação das atividades além de diminuir o custo administrativo relacionado a compra e venda de materiais. Ao otimizar os investimentos em estoques por meio de uma boa administração, tanto em termos de negociações e estratégias de aquisição, quanto de dimensionamento dos estoques, seus custos podem ser significativamente reduzidos gerando elevados ganhos para as empresas (GONÇALVES, 2010).

Gonçalves (2010, p. 4) afirma que:

“uma administração de materiais bem estruturada permite a obtenção de vantagens competitivas por meio da redução de custos, da redução dos investimentos em estoques, das melhorias nas condições de compras mediante negociações com fornecedores e da satisfação de clientes e consumidores em relação aos produtos oferecidos pela empresa”.

Os objetivos de uma gestão de planejamento e controle de estoque são: assegurar o suprimento adequado de matéria prima e material auxiliar; manter o estoque o mais baixo possível mas dentro das margens concebíveis; identificar itens obsoletos ou defeituosos; não permitir a falta ou excesso de produtos em relação a demanda de venda; prevenir-se de perdas, danos, extravios; manter as quantidades em relação aos registros; fornecer bases concretas para a elaboração de planejamentos de curtos, médios e longos prazos; e manter os custos nos níveis mais baixos possíveis (POZO, 2002).

As empresas, de modo geral, preferem manter estoques mínimos para reduzir investimentos, e evitar que o capital da empresa seja imobilizado em materiais. Este capital poderia ser aplicado de outras formas, como na quitação de dívidas no curto

prazo ou até mesmo uma possível aplicação no mercado financeiro. Também são incorporados aos custos de estoque, despesas relacionadas com o armazenamento, que compreende custos para controlar, espaço físico, sistema de armazenagem em movimentação, além do pessoal alocado para este fim. Soma-se a estes fatores, os riscos de perda, roubos, obsolescência ou até mesmo acidentes, como por exemplo, incêndios. Estima-se que o custo total para manutenção dos estoques gira em 25% do valor médio de seus produtos (BALLOU, 2009). Portanto, é importante para a organização manter os custos o mais baixo possível para diminuir os riscos.

A literatura aborda técnicas para gestão e controle de estoques, que tem como principal objetivo a manutenção de níveis baixos levando em consideração a necessidade contínua de disponibilidade de produtos. A seguir, uma breve revisão das principais técnicas é apresentada.

2.1. Análise dos Estoques

A análise de estoque utiliza indicadores que fornecem uma visão integrada ao gestor responsável por suprir a demanda, informando a saúde dos estoques da empresa, para que seja possível tomar decisões utilizando como base as informações disponíveis. Scherr (1989) já afirmava que os custos e a configuração dos estoques devem ser os primeiros itens a serem considerados para a formulação das estratégias relativas a estoques.

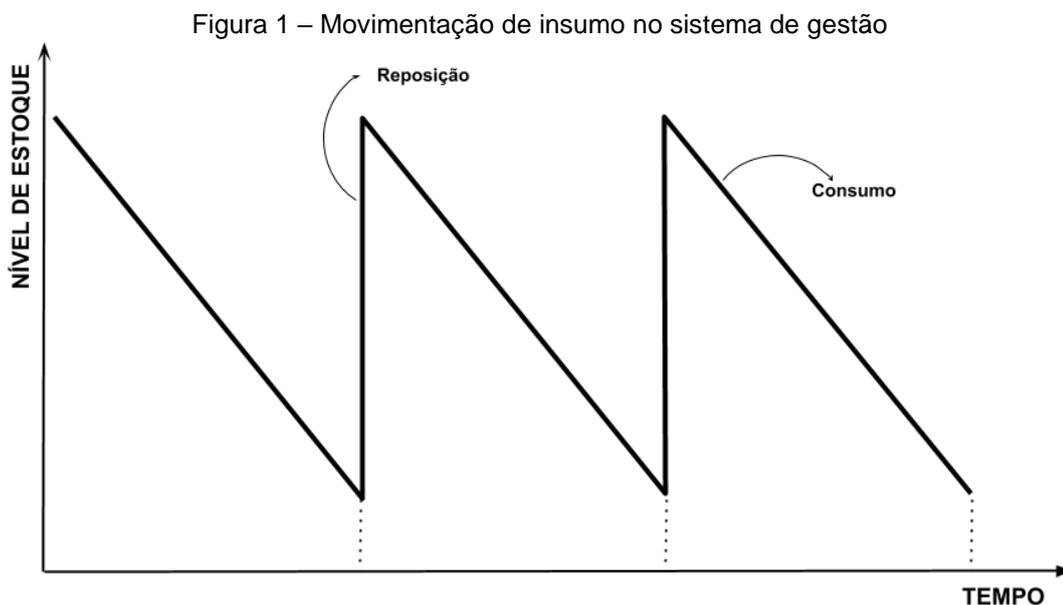
Segundo Martins e Alt (2005, p. 155), “Cabe ao administrador verificar se os estoques estão tendo a utilidade adequada ou sendo um “peso morto”, não apresentando o retorno sobre o capital neles investido”.

Nos anos 80, muitas empresas sofreram consequências negativas ao tentarem buscar a implantação de um sistema comumente chamado de “estoque zero”. Para conseguir chegar a este nível, muitas vezes baixavam os estoques a níveis inferiores às necessidades estratégicas, fragilizando suas operações e permitindo que concorrentes preparados aproveitassem desta falha para realizar ataques competitivos.

Na administração atual, é sabido que os níveis de estoque devem ser reduzidos aos níveis mínimos, porém, tendo em mente que é necessário atender as necessidades estratégicas da organização, o que é improvável de acontecer com os níveis inicialmente imaginados para o “estoque zero”. A gestão desses níveis de estoque é uma das principais funções dos sistemas de administração de produção e não pode ser tratado como uma atividade meramente trivial dentro dos sistemas produtivos das organizações (CORRÊA, GIANESE & CAON; 2007).

Para Corrêa, Gianese & Caon (2007), as principais discussões em torno da gestão de estoques referem-se ao momento e a quantidade de itens a serem adquiridos para reposição dos estoques. Para isso, torna-se necessário definir o momento do ressuprimento e a quantidade a ser reposta para que o estoque seja compatível com as necessidades da demanda. De acordo com Ballou (2009, p. 214), “umas das primeiras questões consideradas na gestão de estoques é a previsão de vendas futuras e a estimativa dos tempos de ressuprimento, desde a colocação do pedido até a chegada do material”.

A Figura 1, demonstra um gráfico dente de serra que é a representação visual da movimentação de um determinado insumo dentro de um sistema de gestão de estoques, sendo que a abscissa demonstra o tempo decorrido para o consumo das unidades. É possível perceber por meio do gráfico que, de tempos em tempos, os níveis de estoque precisam ser restabelecidos.



Fonte: Elaborada pelo autor (2020).

Para Martins e Alt (2005), existem conceitualmente três tipos de modelos de estoques: Modelo de Reposição Contínua (Lote Padrão), Modelo de Reposição Periódica (Intervalo Padrão) e Modelos Híbridos, como o LEC – Lote Econômico de Compra. Tais abordagens são descritas nas próximas subseções.

2.2. Modelo de Reposição Contínua

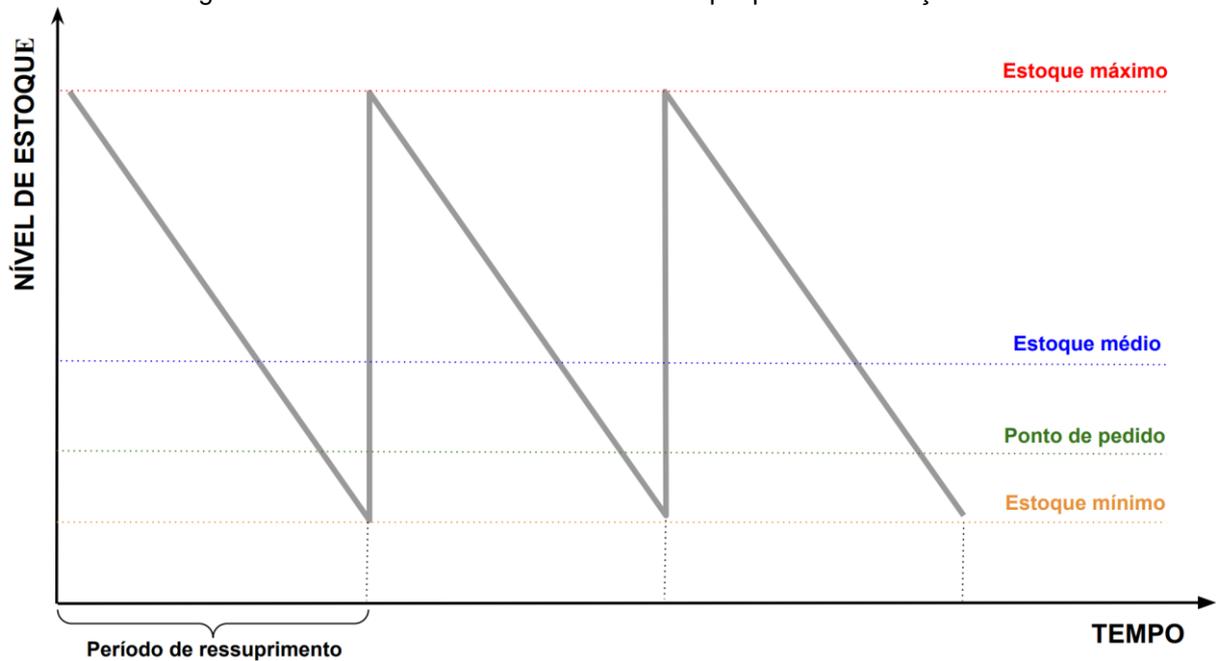
O Modelo de Reposição Contínua consiste no estabelecimento de um nível fixo de reposição, que quando atingido, ativa o gatilho de emissão de um novo pedido com um tamanho previamente definido. Esse nível também é conhecido como Ponto de Pedido (PP). Segundo Buffa (1968), a quantidade a ser suprida pode ser definida utilizando algum critério de interesse baseado na experiência prática do gestor de estoques ou aplicando um modelo de lote econômico que seja apropriado para uma determinada situação.

A adoção do modelo de reposição contínua exige um monitoramento exaustivo do nível de estoque atual para que seja possível identificar o PP. O nível de estoque pode ser entendido como a quantidade de material armazenado para uso da empresa. O estoque máximo indica a capacidade física da empresa de armazenagem, e o estoque médio, representa a soma do estoque inicial – momento em que os níveis se encontram abastecidos – com o estoque final – momento que se torna necessário realizar o abastecimento – dividido por dois.

O PP indica o momento em que a empresa deve tomar ação para repor os níveis de estoque visando o nível identificado como o ideal para a organização. O estoque de segurança, representaria o nível mínimo adequado para atendimento a demanda esperada, enquanto o período de ressuprimento seria a janela temporal para a realização dos pedidos para normalizar os níveis de estoque.

A empresa também deve levar em consideração o tempo de atendimento, que pode ser entendido como o tempo solicitado pelo fornecedor para entregar o pedido da compra. O lote de compra representa a quantidade de material a ser adquirido pela empresa para que o estoque atinja o nível considerado ideal (estoque máximo). A Figura 2 ilustra o modelo de reposição contínua, considerando estes aspectos.

Figura 2 - Monitoramento dos níveis de estoque para identificação do PP



Fonte: Elaborada pelo autor (2020).

O intervalo de tempo entre cada solicitação de compra normalmente é variável, sendo possível definir o momento de realizar o pedido junto ao fornecedor – o PP – por meio da utilização da Equação 1: (ROSA, MAYERLE & GONÇALVES, 2010):

$$r = D \times TR + ES \quad (1)$$

Onde:

- D = número de unidades que se espera que sejam comercializadas (demanda média no tempo);
- TR = tempo médio para ressurgimento; e
- ES = estoque de segurança pré-determinado.

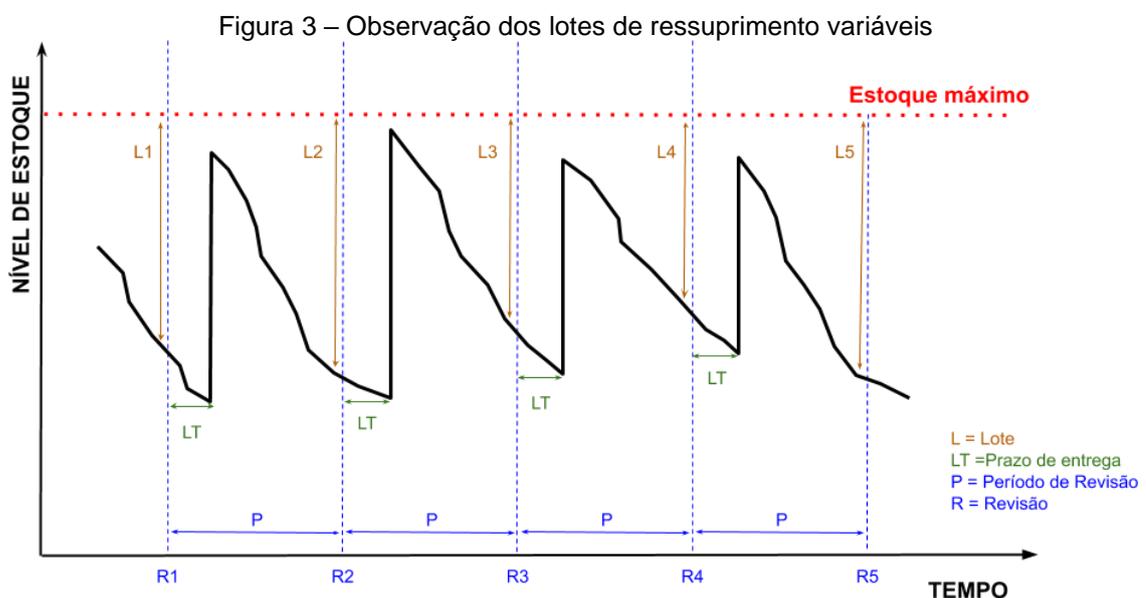
O Modelo de Reposição Contínua pode ser automatizado usando um sistema de informação, na qual seria possível verificar o nível de estoque atual por meio do acompanhamento da saída dos itens comercializados, calcular o tempo de atendimento, e com base nessas informações, o responsável por administrar os estoques pode ser notificado sobre a quantidade exata a ser adquirida em um determinado período para que o nível de estoque seja reabastecido antes que esteja abaixo do nível de segurança e o atendimento a demanda não seja comprometido. A

seguir, será apresentado o Modelo de Reposição Periódica, que utiliza o tempo como “chave” para reabastecimento dos estoques.

2.3. Modelo de Reposição Periódica

O Modelo de Reposição Periódica consiste na realização de uma ordem de compra em intervalos de tempo pré-determinados e com uma quantidade variável. Diferentemente do Modelo de Reposição Contínua, em que os pedidos são realizados com o objetivo de manter sempre o estoque em nível máximo, neste modelo, periodicamente são realizadas verificações dos níveis, e de acordo com o nível de estoque encontrado, é possível determinar a quantidade a ser reposta para que seja atingido um nível previamente planejado (CORRÊA, GIANESI & CAON, 2007).

Como a demanda é uma variável aleatória e sofre variações nos intervalos observados durante as revisões de estoque no modelo proposto por esse sistema, a quantidade de compra (QC) é variável, tornando a quantidade do pedido em cada revisão diferente em relação a demanda, porém, a única constante no modelo é o período entre as revisões de estoque (GONÇALVES, 2010). A Figura 3 ilustra o sistema de revisão periódica, trazendo a observação do lote total, os lotes adquiridos para reposição (L1, L2, L3, L4 e L5), os períodos de revisão (P), o número de revisões realizadas no tempo (R1, R2, R3, R4 e R5) e o prazo de entrega ou *lead time* (LT).



Fonte: Adaptada de Lourenção *et al.* (2017).

Para determinar a quantidade do item a ser adquirida (QC), é possível utilizar a Equação 2 (LOURENÇÃO *et al.*, 2017):

$$QC = EM - (EP + QP) \quad (2)$$

Onde:

- QC = quantidade a ser comprada;
- EM = estoque máximo, limitado pela capacidade de armazenamento;
- EP = estoque presente; e
- QP = eventual quantidade já adquirida.

Para determinar o estoque máximo, deve-se observar as revisões de estoques conduzidas em cada período, considerando o tempo adicional para que o processo de reposição seja concluído. Gonçalves (2010, p. 159) conclui que “o estoque máximo deverá ser suficiente para atender à defasagem de tempo entre as revisões e, além disso, suportar o atendimento até que uma nova encomenda seja entregue”.

A seguir, a técnica de Lote Econômica de Compra (LEC) é apresentada.

2.4. Lote Econômico de Compra (LEC)

O lote econômico de compra (LEC) tem como objetivo, obter a melhor quantidade a ser adquirida em função dos custos de armazenamento e manutenção do estoque frente a realização do pedido (HARRIS, 1913). Em outras palavras, o LEC permite encontrar o volume ideal de itens a serem comprados e estocados para minimizar o custo total do estoque. Este modelo busca a melhor estratégia para que seja possível encontrar a melhor quantidade para se manter em estoque e em qual periodicidade um novo pedido deverá ser realizado.

Para encontrar o LEC, tomaremos como exemplo uma empresa que venda 1.000 unidades de um determinado item mensalmente. Considere que o custo de estocagem por produto é de R\$ 1,00 ao mês e o custo para ressurgimento deste item seja de R\$ 45,00 por unidade. A Equação 3 permitirá encontrar o lote econômico de compra deste item:

$$LEC = \sqrt{\frac{(2 \times D \times CpU)}{CA}} \quad (3)$$

Onde:

- D representa o número de unidades que se espera que sejam comercializadas (demanda média no tempo);
- CpU representa o custo por unidade do produto para reposição; e
- CA representa o custo de armazenagem do produto mensalmente.

Aplicando a Equação 3 para encontrar o LEC para o exemplo acima, poderemos encontrar o seguinte resultado:

$$LEC = \sqrt{\frac{2 \times 1000 \times 45}{1}} = 300 \text{ unidades} \quad (3.1)$$

Neste caso, concluímos que o pedido para repor o estoque com a quantidade ideal seria 300 unidades. O número de pedidos por mês será 3,33 ($NP = 1.000/300$) e o tempo para cada pedido é de 9 dias ($30 \text{ dias}/3,33 \text{ pedidos}$). Levando em consideração este cenário, no momento inicial ($t = 0$) a empresa está com seu nível de estoque zerado e receberá 300 unidades do item, que são consumidas em sua totalidade em 9 dias. Quando isso ocorrer, o estoque deverá ser repostado pelas 300 unidades adquiridas no primeiro momento. Quando $t = 18$ (9 dias após a primeira reposição) os estoques chegarão a zero novamente, sendo um novo pedido de 300 unidades realizado para ressuprimento. O processo será repetido uma vez a cada 9 dias.

Os custos podem ser resumidos, utilizando as Equações 4, 5 e 6. (ROGERS, RIBEIRO & ROGERS, 2004, p. 7-8):

$$CTP = CpU \times NP = 45 \times 3,33 \dots = R\$ 150 \quad (4)$$

$$CAT = CA \times \left(\frac{LEC}{2}\right) = 1 \times \left(\frac{300}{2}\right) = R\$ 150 \quad (5)$$

$$CT = CTP + CAT = R\$ 300 \quad (6)$$

Onde:

- CTP representa o custo total do pedido, considerando todos os itens que o compõem;

- *CAT* representa o custo de armazenagem total; e
- *CT* representa o custo total.

Desta forma, podemos concluir que nenhum outro lote de compra minimiza os custos totais se a quantidade a ser reposta por meio de um pedido seja maior ou menor a 300 unidades, por conta dos custos de armazenagem. Nesse ponto, o custo do pedido e da armazenagem se igualam, sendo dessa forma, otimizados. Qualquer pedido acima deste ponto, tornará o custo de armazenagem alto, inviabilizando a otimização sobre a ótica de custos.

Os modelos clássicos de gestão de estoques não levam em consideração os dados históricos referentes às vendas, que também influenciam na quantidade de itens a serem comprados para reposição e a época em que determinados insumos precisam ser adquiridos em maior ou menor quantidade. Na verdade, uma das hipóteses do modelo mais tradicional da gestão de estoques, o LEC, emprega os maiores esforços nos custos operacionais e financeiros envolvidos, visando minimizar os custos logísticos como um todo, mas sem considerar os aspectos sazonais. Portanto, possíveis variações na demanda como promoções, liquidações, datas festivas e outros eventos que possam influenciar na demanda, são desconsiderados.

A seguir, serão apresentadas algumas técnicas de predição de demandas que auxiliam na gestão de estoques nas organizações.

3. TÉCNICAS DE PREDIÇÃO DE DEMANDAS

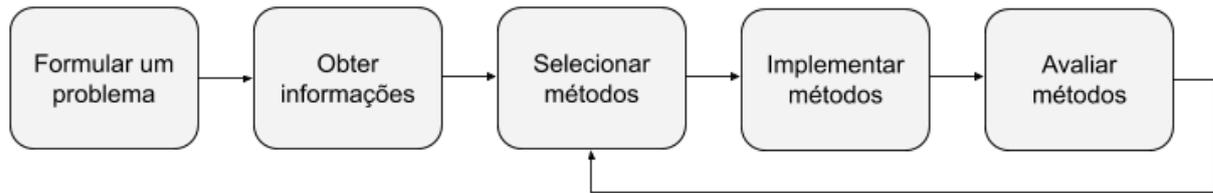
De acordo com Pearce (1974), a razão fundamental para a necessidade de realizar tarefas de predição é em consequência dos problemas gerados pelas organizações e porque a situação futura é sempre incerta para o tomador de decisões. Se tratando de administração e atividades de gestão, existe uma grande necessidade de realizar atividades de planejamento porque o tempo de espera para a tomada de decisão varia de alguns anos (como por exemplo, investimentos a longo prazo na bolsa de valores), alguns dias ou horas (no caso de agenda de produção ou logística) ou até mesmo de alguns segundos (como por exemplo, para rotas de telecomunicação ou balanceamento de cargas elétricas).

Para Lancaster & Lomas (1985), a tarefa dos gestores de empresas seria mais simples se as indústrias não vivessem um estado constante de mudança. Essas mudanças causam impacto no nível de competição, as atividades das empresas se tornam mais complexas e convergem para a internacionalização entre os negócios e as circunstâncias, de modo geral, tendem a se modificar muito mais rápido. A predição pode ser vista como uma forma de dar a vantagem por meio da informação prévia ao gestor, para que este possa tomar uma ação proativa para explorar a oportunidade. Além de ser uma ferramenta importante para realização de planejamentos eficientes e eficazes (MAKRIDAKIS, WHEELWRIGHT & HYNDMAN, 1997).

Para Armstrong (2002), os tomadores de decisão precisam de previsões somente em cenários em que há incerteza sobre o futuro. A grande maioria das empresas convivem diariamente com incertezas sobre o futuro, especialmente em se tratando do mercado.

Por meio de técnicas, as organizações são capazes de entender e dominar certos padrões de consumo, e a partir destas descobertas, realizar previsões para agir de forma proativa. Esta atividade pode ajudar pessoas e organizações a planejar o futuro e tomar decisões racionais. A Figura 4 exemplifica o processo de utilização de predições em empresas.

Figura 4 - Processo de utilização de predições



Fonte: Adaptada de Armstrong (2002).

Existem cinco passos básicos em qualquer tarefa de predição quando dados quantitativos estão disponíveis (ARMSTRONG, 2002), que são:

1. Formular um problema: Geralmente a etapa mais difícil para o responsável pela predição, pois envolve um profundo entendimento sobre a utilidade da predição, quem a irá utilizar, a quais objetivos estratégicos da empresa a predição está ligada, e qual o problema que pretende se resolver com esta tarefa.
2. Obter Informações: A maioria das organizações possuem ao menos dois tipos de informações disponíveis: dados estatísticos (números) e o conhecimento e julgamento de especialistas do setor. Os dois tipos de informação devem ser utilizados ao realizar predições. Portanto, precisamos coletar os dados históricos sobre os itens de interesse de modo que sirvam na construção do modelo a ser utilizado.
3. Selecionar métodos: Uma vez que o problema esteja descrito e os dados disponíveis, um modelo teórico deve ser desenvolvido para responder o questionamento que originou a predição. Se, por exemplo, o problema está relacionado com as vendas, produção ou utilização de matéria prima e a empresa dispõe dos dados para estes itens, um modelo quantitativo seria mais apropriado. Para outras situações em que os dados históricos não são numéricos ou a experiência do cliente, por exemplo, for mais importante, dados qualitativos podem ser empregados para realizar predições.
4. Implementar métodos: Esta etapa trata da geração do modelo para realização das predições uma vez que os dados foram obtidos, preparados e um modelo foi selecionado. Dados referentes a períodos recentes são guardados e posteriormente utilizados para verificação da eficiência do modelo desenvolvido.

5. Avaliar métodos: Neste momento, os dados gerados pelo modelo preditivo são comparados com dados reais após a implementação do modelo escolhido. Os erros encontrados na predição, são analisados e busca-se formas de diminuir estas margens para o menor valor possível. Com essas correções, o modelo é retroalimentado até que a abordagem ideal seja alcançada.

Sethi, Yan & Zhang (2005) afirmam que empresas precisam lidar com clientes que possuem demandas em diferentes graus de variação e para atender essas demandas, é necessário desenvolver a habilidade de predição. Os autores afirmam que empresas com habilidades superiores de predição podem se dar ao luxo de produzir grande parte de sua demanda fazendo uso de modos de produção mais lentos e com menos gastos em serviços logísticos, pagando mais para produções mais rápidas e por serviços logísticos apenas quando a demanda surge de maneira inesperada. Portanto, empresas com demandas irregulares e inabilidade de realizar predições precisam arcar com custos altos para responder aos aumentos de demanda inesperados.

Segundo Doganis *et al.* (2006), para vencer a forte competição atual, a maioria das organizações emprega esforços contínuos para aumentar seus lucros e reduzir seus custos. Para os autores, uma predição de demanda confiável é certamente um caminho mais “barato” (sobre a ótica de custos) de atingir estes objetivos, uma vez que as atividades de predição melhoram os resultados da organização por meio da evolução no atendimento ao cliente, reduzir o custo de vendas perdidas e devoluções é uma forma de planejar mais eficientemente as operações. Especialmente para a indústria alimentícia, um modelo de predição de demanda pode ser benéfico por conta da vida curta dos produtos perecíveis e a importância de produtos de qualidade para a saúde humana.

Apesar de as técnicas de predição contribuírem de forma significativa na tarefa de gestão de estoques, não significa que estas não possuem falhas. Para Novaes (2007), a previsão da demanda depende diretamente de desdobramentos futuros intrinsecamente conectados com aspectos políticos, sociais e econômicos, tanto locais quanto estrangeiros e, portanto, está sujeita a erros. O responsável por elaborar a predição da demanda deve manter-se atento para alguns pontos a serem observados:

- As previsões estão sujeitas a erros;
- As projeções de longo prazo geralmente são menos confiáveis do que as de curto prazo. Como as mudanças de contexto (econômicas, políticas etc.) ocorrem em um ritmo mais lento, é improvável que sofram alterações bruscas a ponto de afetarem previsões de curto prazo. Já as mudanças sazonais, como as trocas de estações, por exemplo, são previsíveis dentro dos limites nos quais se manifestam.
- Quanto mais dados forem agregados, mais precisas serão as projeções, desde que a agregação seja feita de forma lógica. Um exemplo a ser citado seria a projeção da venda de automóveis, que tenha dados agregados ano a ano para todo o Brasil. Neste caso, a projeção apresentará um erro relativo menor do que se a mesma previsão fosse realizada estado a estado, de maneira separada.

De acordo com Ross (2015), existem sete características ou dimensões que tem papel determinante em uma predição efetiva:

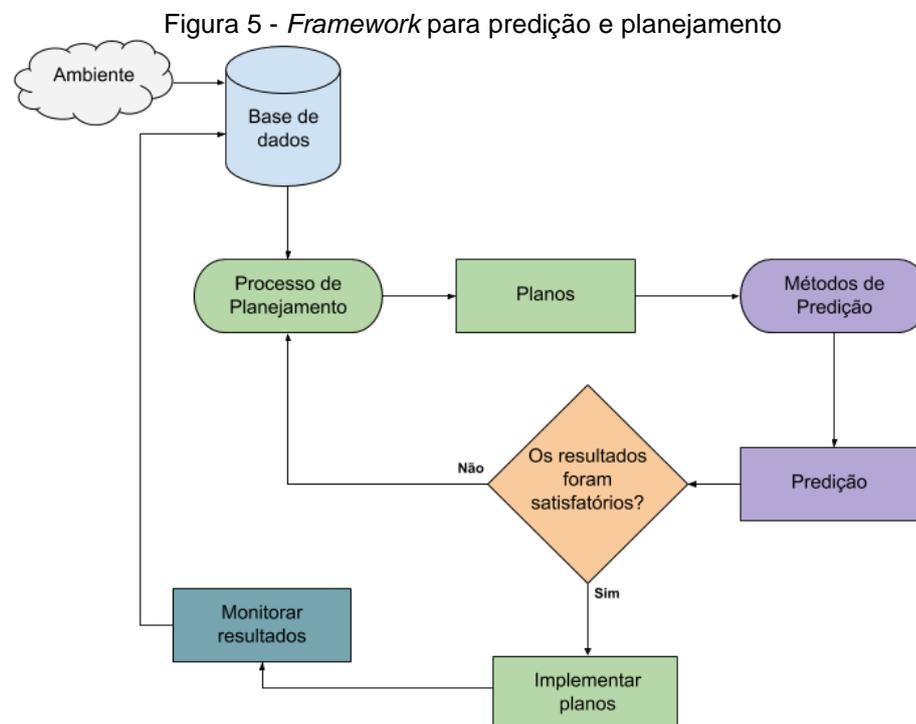
1. Horizonte temporal: A quantia de tempo observada influencia na escolha do modelo de predição. Cada nível de predição possui características únicas relativas ao modelo de predição mais apropriado e a quantidade dos dados necessários. Projeções de longo prazo geralmente utilizam modelos qualitativos enquanto modelos quantitativos são mais apropriados para projeções de curto prazo.
2. Nível de detalhe dos dados: Para realizar uma projeção, é necessário identificar dois aspectos: a técnica empregada e o nível de detalhe necessário. Ao selecionar uma predição específica para um plano, o responsável precisa determinar o nível de detalhe necessário para que projeção se torne útil na tomada de decisão. As projeções podem ser feitas em alto nível, por exemplo, a quantidade total de itens a serem vendidos em um determinado horizonte temporal ou podem ser feitas mais específicas, detalhando a venda por tipo de itens e o horário mais provável em que estes itens sejam comercializados.
3. Tamanho da base de dados disponível: O número de elementos em uma base de dados a ser utilizada para predição tem um impacto direto nos métodos a serem adotados. À medida que os dados históricos se

acumulem sobre uma determinada informação, mais simples será o método de predição a ser adotado, por conta do emprego de métodos estatísticos. Por este motivo, quanto menor a amostra de dados disponível, técnicas mais complexas deverão ser aplicadas para diminuir a variação e tornar o resultado mais confiável.

4. Controle de previsão: Os responsáveis pelo planejamento precisam utilizar métodos que gerem *feedbacks* confiáveis e em tempo para que seja possível fazer um controle da previsão. Os métodos de controle precisam indicar quando os dados extrapolaram os limites de variância previamente definidos e precisam ser capazes de se ajustar a mudanças nos padrões do histórico de demandas. O principal objetivo é assegurar que as decisões relativas a previsões sejam tomadas de acordo com eventos atuais.
5. Constância: Os modelos de predição utilizados para projetar a demanda que demonstram estabilidade com o passar do tempo, se diferem de modelos que precisam ser ajustados regularmente. Em ambientes estáveis, uma predição baseada em dados históricos pode ser adotada e revisada periodicamente para verificar se ainda é apropriada para uma determinada realidade. Por outro lado, quando a variação dos dados previstos for muito grande, os responsáveis pela predição deverão aplicar métodos para adaptar a previsão constantemente para refletir na realidade atual vivida pela empresa.
6. Seleção do modelo de predição: Os modelos devem ser selecionados de acordo com a dimensão relevante para a predição. Ao planejar o atendimento a uma demanda, é necessário entender quais dimensões são relevantes para a predição: área geográfica, grupos de consumidores, faixa etária, entre outros. É importante entender as diferenças na demanda entre essas dimensões e utilizar modelos compatíveis com os padrões de consumo avaliados.
7. Alinhamento com os processos de planejamento: Métodos para predição de demandas devem ser escolhidos para fortalecer o planejamento e o processo de tomada de decisões da empresa, além de melhorar as operações. No entanto, é comum observar a confusão que algumas empresas fazem entre predição e planejamento. Os

responsáveis por realizar as previsões, geralmente empregam métodos para simular os resultados dos planos desenvolvidos. Depois da validação dos resultados, os responsáveis revisam os planos, fazem uma nova previsão e repetem este processo até que os resultados das simulações se mostrem satisfatórios.

Para Armstrong (2002), a previsão é geralmente confundida com planejamento. O planejamento diz respeito a um cenário no sentido de “como ele deveria ser”, enquanto a previsão diz respeito a um cenário no sentido de “como ele será”. Planejadores podem utilizar métodos de previsão para projetar os resultados de planos alternativos. Se caso o resultado não seja satisfatório, os planos podem ser revisados e uma nova previsão realizada, repetindo o processo até que o resultado atenda as expectativas. Os resultados do planejamento podem ser monitorados e serem utilizados no próximo ciclo de planejamento. No contexto deste trabalho, o planejamento se aplica a gestão de estoques da organização objeto do estudo. A Figura 5 ilustra um framework, sugerido por Armstrong (2002), para previsão e planejamento.



Fonte: Adaptada de Armstrong (2002).

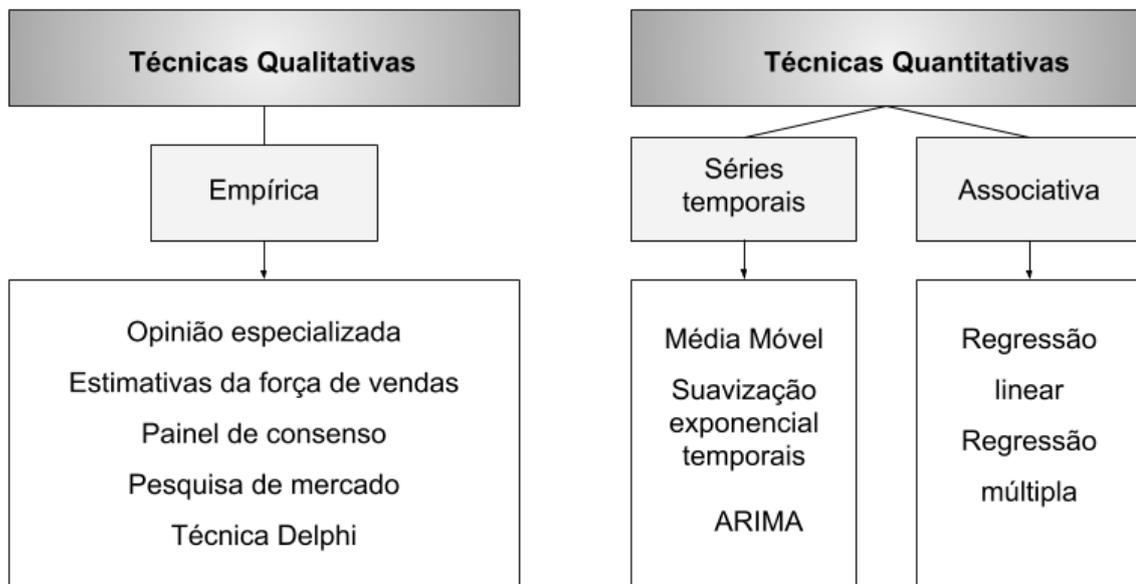
Para Armstrong (2002), a previsão atende várias necessidades, podendo ajudar pessoas e organizações a planejar o futuro e tomar decisões racionais. Se

utilizarmos como exemplo a empresa objeto deste estudo, podemos utilizar métodos de predição para provisionar a quantidade de estoque necessária para atender a demanda, utilizando os dados históricos acumulados nos sistemas de informação que fazem a gestão de vendas da rede.

Novaes (2007, p. 165), afirma que “a escolha do método mais apropriado para se fazer uma determinada previsão da demanda vai depender da análise criteriosa dos requisitos listados anteriormente, vis-à-vis a situação real do problema”. O autor observa que existem algumas ferramentas (softwares) no mercado que possuem rotinas de previsão embutidas, ao exemplo da SAP¹

Para Ross (2015), a variedade de técnicas de predição pode ser organizada em três categorias básicas: qualitativa, quantitativa e associativa. Cada categoria possui diversas técnicas que podem ser utilizadas. A Figura 6 demonstra alguns tipos de técnicas de predição de acordo com sua categoria.

Figura 6 - Categorias de técnicas de predição



Fonte: Adaptada de Ross (2015).

A primeira usa basicamente métodos não-estatísticos como a opinião especializada, intuição e avaliação subjetiva para determinar previsões. Os modelos

¹ Sigla de *Systeme, Anwendungen und Produkte in der Datenverarbeitung*, que é traduzida para o português como Sistemas, Aplicativos e Produtos para Processamento de Dados, é o nome da empresa alemã criadora de software para gestão de empresas de mesmo nome, cujo site oficial está disponível em: <https://www.sap.com/index.html>

quantitativos aplicam modelos estatísticos, análises de séries temporais e projeções em busca de padrões históricos que possam ser extrapolados para o futuro. Já a associativa, tem como objetivo expressar matematicamente as relações entre o objeto da predição e outras variáveis como eventos socioeconômicos, políticos, tecnológicos, entre outros, que podem influenciar a predição. É importante observar que o processo de predição é um modelo iterativo e incremental, como visto na Figura 5.

Conforme os fatos se desenvolvem, o modelo preditivo deve ser capaz de fornecer aos gestores, a capacidade de processar as informações geradas a partir dos dados disponíveis. Geralmente, são utilizados softwares ou linguagens de programação para realizar os cálculos de maneira eficiente e precisa, permitindo que os responsáveis pela predição possam agir rapidamente para corrigir erros no modelo, avaliar as variâncias e realizar alterações para melhorar o grau de confiança do modelo.

Os métodos qualitativos podem trazer ganhos significativos, pois agregam diferenciais percebidos por especialistas no setor ou pessoas que tenham experiência em determinado mercado ou segmento. Quando utilizados em conjunto com métodos quantitativos, podem enriquecer o processo decisório, pois as informações encontradas podem ser interpretadas por meio das técnicas qualitativas. Na Seção 3.1, algumas das técnicas qualitativas geralmente utilizadas serão apresentadas.

3.1. Técnicas qualitativas

Para Moon (2013), pode ser definida como o processo de acumular conhecimento, opiniões e intuições de pessoas com experiência, tornando estes elementos em previsões formais. É o processo explícito de tentar visualizar o futuro incorporando todas as informações disponíveis para prever a demanda. Ainda segundo o autor, pode ser necessário utilizar a previsão de demanda qualitativa, quando existem razões para crer que o futuro não será necessariamente como o passado, ou seja, existem fatores internos ou externos que podem alterar a demanda futura. Essas informações não são visíveis se os dados forem analisados estritamente sob o aspecto quantitativo.

De acordo com Ross (2015) as técnicas qualitativas são geralmente utilizadas para desenvolver previsões em um nível estratégico. O seu principal objetivo é usar o julgamento humano baseado em análises dos dados disponíveis, percepção de fatores externos ou analogia histórica para desenvolver uma previsão de demanda.

Vu (2015) classifica as técnicas qualitativas em três categorias primárias: julgamento individual, painel de especialistas e a técnica Delphi.

1. Julgamento individual: é um modelo preditivo criado por um especialista em um certo campo de atuação que se baseia na sua experiência pessoal, suas conquistas no mercado e o estado atual do negócio e da economia de modo geral. O especialista pode empregar algumas técnicas, como a analogia comparando itens similares, a análise de cenários para antecipar um acontecimento futuro de curto prazo e coletando dados qualitativos por meio de pesquisas e formulários, compilando esses dados e realizando uma análise qualitativa dos resultados.
2. Painel de especialistas: criado por um grupo de profissionais em que suas opiniões são combinadas, pesadas e ajustadas durante discussões e avaliações entre todos os membros do grupo. Um participante pode mediar as discussões com o objetivo de gerar um consenso no final do painel. É uma espécie de combinação entre vários julgamentos individuais.
3. Técnica Delphi: este método é similar ao painel de especialistas, mas os membros participantes não podem discutir o problema uns com os outros. Para contornar essa situação, um mediador aplica uma série de questões que são respondidas de forma anônima e são compiladas posteriormente. Com base nas respostas anteriores, um novo questionário é gerado e o processo se repete até que o consenso seja atingido.

Ross (2015) acrescenta as seguintes categorias complementares:

4. Estimativa da força de vendas: usada para projetar vendas futuras compiladas pela força de vendas da empresa, usando como base as estimativas individuais de cada vendedor da empresa. A vantagem de utilizar essa técnica é o emprego do conhecimento especializado das pessoas que atuam diretamente com as vendas para o consumidor final.

5. Pesquisa de mercado: esta técnica procura prever a demanda futura e suas tendências aplicando questionários diretamente aos consumidores de um determinado segmento, buscando entender seus padrões de comportamento que indicariam um possível padrão de consumo futuro.

Lomas & Lancaster (1985) apontam que o maior problema relacionado com técnicas subjetivas se comparadas com as objetivas é o grau de incerteza, uma vez que a predição também é subjetiva. Podem existir duas ou mais predições similares, criadas por pessoas diferentes, e cada uma apresentar um grau de certeza diferente, uma vez que o julgamento de cada um é influenciado pela forma de interpretação e vieses.

Para Lomas & Lancaster (1985, p. 143), “técnicas subjetivas dependem mais de opinião e menos de matemática em sua computação”². Portanto, segundo os autores, o principal risco de usar as técnicas qualitativas é que seu resultado algumas vezes parece científico e em consequência, acaba recebendo mais créditos do que deveria. A realidade demonstra que, na maioria dos casos trata-se apenas de “palpites educados”, por se tratarem muitas vezes de julgamentos baseados em subjetividade, sem que sejam explicitados critérios objetivos ou modelos matemáticos que explicam seu resultado. Neste caso, a qualidade do modelo fica condicionada ao conhecimento e experiência do responsável por realizar as predições. É importante salientar que as técnicas qualitativas são insuficientes quando utilizadas de forma isolada, mas podem fortalecer as técnicas quantitativas por trazerem elementos que denotam experiência e/ou opiniões especializadas sobre determinado assunto. Na próxima seção, são detalhadas algumas técnicas quantitativas usadas na obtenção de modelos replicáveis.

² Tradução livre do trecho original: “*subjective techniques depend more opinion and less on mathematics in their computation*”.

3.2. Técnicas quantitativas

Os modelos quantitativos, também conhecidos como modelos estatísticos, são abordagens objetivas para a atividade de predição e dominam este campo porque provêm uma série de passos sistemáticos que podem ser replicados (HOSHMAND, 2010). A abordagem quantitativa inclui técnicas como as séries temporais e as análises associativas. As séries temporais fazem predições de valores de uma variável se baseando no seu comportamento histórico em um horizonte temporal enquanto as abordagens associativas, são baseadas na investigação de diversos fatores externos que podem afetar as tendências de mercado. Essa abordagem utiliza técnicas de regressão ou não-regressão, que incluem a média móvel, suavização exponencial, decomposição em séries temporais e Box-Jenkins (ROSS, 2015).

Segundo Novaes (2007, p. 168), “o nível de demanda apresenta tendência quando seu valor for mudando sistematicamente ao longo do tempo”. Para o autor, quando uma previsão demonstra tendência, geralmente utiliza-se um modelo de regressão, que se tornou um processo estatístico amplamente conhecido e utilizado por empresas. Quando a demanda é dependente apenas de uma variável, como a temperatura, pode-se utilizar a regressão simples. No caso de existirem duas ou mais variáveis independentes, utiliza-se a regressão múltipla. Maiores detalhes sobre séries temporais serão apresentados na próxima seção.

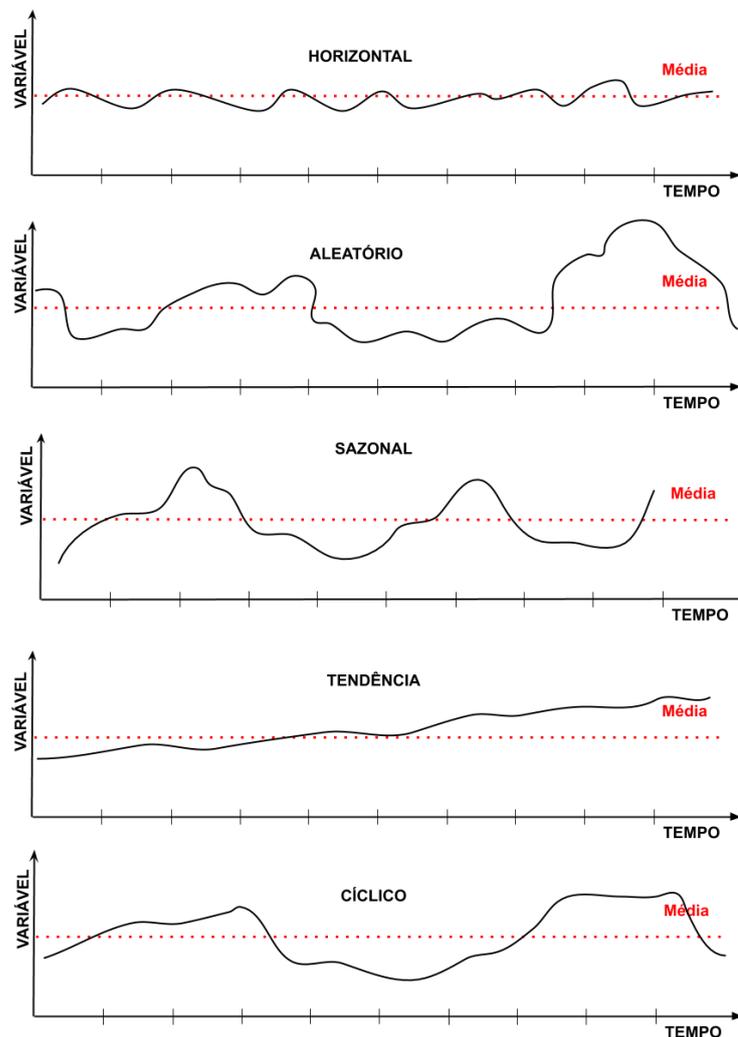
3.2.1. Séries Temporais

Séries temporais são um conjunto de observações geradas sequencialmente no decorrer de um período (Box *et al.*, 2015). Qualquer variável que construa dados que são coletados, armazenados e observados em sucessivos incrementos temporais é chamada de série temporal. As séries temporais observam medidas, fatos e unidade tempo. A quantidade de cerveja produzida mensalmente em um país é um exemplo de séries temporais (HANKE & WICHERN, 2014). A revisão de dados históricos durante um período provê aos gestores e aos tomadores de decisão, uma visão geral sobre acontecimentos passados, que podem ser utilizados para obter estimativas relacionadas ao desempenho futuro. Para Ross (2015), uma vez que as estimativas

de demandas são incertas, um bom método de previsão se torna um fator crítico de sucesso para uma empresa.

Para Henke & Wichen (2014), uma das etapas mais importantes para a seleção de métodos de previsão para dados com séries temporais é considerar os diferentes tipos de padrão de dados, que podem ser divididos em cinco categoriais gerais: horizontal, aleatório, tendência, sazonal e cíclica. A Figura 7 ilustra o comportamento da variável de interesse no tempo de acordo com o tipo de variação.

Figura 7 - Tipos de séries temporais



Fonte: Adaptada de Ross (2015).

Quando os dados são coletados conforme passa o tempo e flutua em torno de um nível constante ou médio, trata-se de um padrão horizontal de dados. Vendas mensais de um produto que não aumentam ou diminuem durante um período pode ser considerado um padrão horizontal de dados. Para Ross (2015), a variação

horizontal exibe padrões que são relativamente estáveis e consistentes, com pequenas variações na demanda no decorrer do tempo. Esses padrões são características de itens com vendas estáveis.

Para Ross (2015), a série temporal com variação aleatória, não possui padrão histórico visível. Demandas aleatórias geralmente enfrentam grandes variações e são imprevisíveis, pois representam uma demanda dinâmica.

De acordo com Hoshmand (2010), a variação sazonal se refere a flutuações repetitivas que acontece no período de um ano. Um padrão repetitivo de variação sazonal amplamente conhecido, por exemplo, é o clima. Um outro padrão de variação sazonal, seria o aumento da venda de bebidas quentes em cafeterias durante o inverno. São variações que tem um horizonte temporal conhecido e se repetem ano a ano.

Quando um dado demonstra sinais de subida ou descida durante um período, trata-se de um padrão de tendência dos dados. O aumento do preço médio do aluguel na região central de uma cidade por conta do aumento populacional, pode demonstrar um padrão de tendência por exemplo (ROSS, 2015). Para o autor, as variações cíclicas, são movimento ondulares que são observados durante períodos estendidos de tempos. O impulso econômico causado por conta da revolução industrial, pode ser caracterizado como um ciclo que durou por décadas. As variações cíclicas são difíceis de serem previstas e podem ser resultados de inúmeras causas distintas.

Na variação sazonal, envolve séries temporais que exibem padrões repetidos em períodos específicos, aumentando ou diminuindo a demanda de acordo com o período do ano. Na série temporal com tendência, pode-se observar uma subida ou declínio constante, de acordo com o passar do tempo. Uma tendência existe quando o padrão da demanda se mantém em subida ou declínio por aproximadamente sete ou mais períodos. E por último, no tipo de série temporal cíclica, é observada uma curva acentuada de subidas e descidas, por longos períodos.

3.2.2. Média Móvel

Um dos problemas frequentemente encontrados pelos gestores é preparar previsões de curto prazo para muitos itens. O custo de desenvolver modelos para previsão de cada um dos itens é impraticável, especialmente se avaliados pela ótica de custos e de tempo, logo, é necessário escolher uma técnica que poderá ser empregada de maneira fácil e descomplicada para cada um dos itens e que também proverá previsões com um razoável grau de confiança no curto prazo nas quais elas são requisitadas (LANCASTER & LOMAS, 1985).

Para Vu (2015), a média móvel é o modelo mais simples e fácil de ser aplicado, tratando-se de séries temporais. Se, por exemplo, quisermos fazer uma previsão de quantos carros de uma determinada marca e categoria serão vendidos no próximo trimestre, podemos nos basear no desempenho passado (dados históricos) da loja em questão. Esta técnica vai assumir que a previsão seria uma média entre o que ocorreu no passado e o que está acontecendo no presente. Um modelo simples de média móvel de um valor x é descrito pela Equação 7:

$$P_{t+1} = MM(x)_t = \frac{\sum_{i=t-x+1}^n A_i}{x} \quad (7)$$

Onde:

- P_{t+1} = previsão do valor para o período $(t + 1)$;
- A_i = o valor atual para o período t ;
- x = quantidade de médias móveis;
- n = ordem das médias móveis;
- t = período no presente.

Um exemplo para ilustrar a utilização da técnica, seria a previsão de faturamento de uma empresa para o segundo trimestre do ano, utilizando os dados do primeiro trimestre como referência. Vejamos na Tabela 1, a demonstração dos valores:

Tabela 1 - Faturamento do segundo trimestre de uma empresa

| Mês | Faturamento |
|------------|--------------------|
| Abril | R\$ 42.000,00 |
| Maio | R\$ 38.000,00 |
| Junho | R\$ 40.000,00 |

Fonte: Elaborada pelo autor (2020).

Aplicando a Equação 7 encontraremos a seguinte previsão:

$$P_{Abril} = P_4 = MM(3)_3 = \frac{(42.000+38.000+40.000)}{3} = R\$ 40.000,00 \quad (7.1)$$

$$P_{Maio} = P_5 = MM(3)_4 = \frac{(38.000+40.000+40.000)}{3} = R\$ 39.333,00 \quad (7.2)$$

$$P_{Junho} = P_6 = MM(3)_5 = \frac{(40.000+40.000+39.333)}{3} = R\$ 39.777,00 \quad (7.3)$$

Os resultados foram arredondados para simplificar os valores, e a Tabela 2 ilustra a previsão encontrada por meio da aplicação da fórmula.

Tabela 2 – Previsão para o terceiro trimestre de uma empresa

| Mês | Faturamento |
|------------|--------------------|
| Julho | R\$ 40.000,00 |
| Agosto | R\$ 39.333,00 |
| Setembro | R\$ 39.777,00 |

Fonte: Elaborada pelo autor (2020).

É importante salientar, que caso a periodicidade seja par (MM = 6, por exemplo), podemos encontrar uma situação de assimetria de valores, no entanto, com periodicidades mais altas o erro ocasionado pela assimetria é pouco significativo (NOVAES, 2007). Lancaster & Lomas (1985), afirmam que a técnica das médias móveis não responde bem a mudanças bruscas em vendas. Se um “ponto de virada” se manifesta (como por exemplo, a tendência está para uma alta em vendas, mas na verdade muda para uma queda), levará apenas um ou dois períodos de predição para que essa mudança seja refletida na previsão realizada. Esse método é lento em relação as mudanças porque é atribuído um peso uniforme para todos os valores que compõem a predição. Na próxima seção, é apresentada a técnica de Suavização Exponencial, que pode contribuir nessas situações.

3.2.3. Suavização Exponencial

Para Novaes (2007), por conta da volatilidade observada no mercado consumidor, sendo este sujeito a oscilações econômicas, impactos políticos internos e externos, observa-se a necessidade da utilização de métodos de previsão mais dinâmicos, que sejam capazes de relevar essas mudanças caso elas se manifestem. Um dos métodos para realizar este tipo de previsão é a suavização exponencial, que permite que o responsável pelas previsões atribua pesos aos dados históricos e informações recentes, com o objetivo de refletir possíveis variáveis que possam alterar o curso das demandas, incluindo tendência e sazonalidade. O processo é dinâmico, pois é facilmente ajustável para refletir mudanças observadas no mercado no momento de construir a previsão.

A fórmula matemática para realizar essa suavização é dada pela Equação 8 (MOON, 2013):

$$P_{t+1} = \alpha D_t + (1 - \alpha)P_t \quad (8)$$

Considerando o fator α (alfa), valor expresso entre 0 e 1:

$$0 < \alpha < 1$$

Onde:

- P_{t+1} = previsão a ser realizada;
- t = período;
- D_t = demanda; e
- α = ponderação a ser considerada.

Vejamos um exemplo: se em uma empresa de veículos a previsão atual de vendas de um determinado modelo é de 1.000 unidades, e a demanda atual é de 900 unidades, digamos que o fator alfa definido pela responsável pela previsão seja de 0.5, então a suavização exponencial pode ser aplicada utilizando a Equação 8:

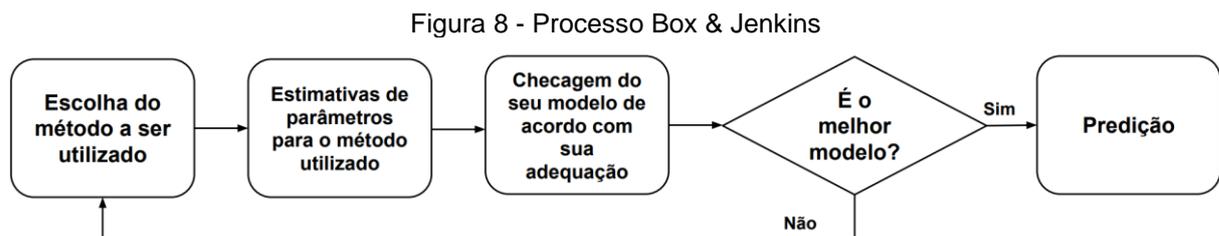
$$0.5 \times 900 + (1 - 0.5) \times 1.000 = 950 \text{ unidades} \quad (8.1)$$

Para Lancaster & Lomas (1985) não existe uma regra específica para determinar o valor de α , sendo a prática mais comum experimentar valores utilizando

dois ou três valores diferentes aos dados históricos e verificar quais destes valores produzem o menor intervalo de erros.

3.2.4. ARIMA

O trabalho proposto por Box & Jenkins (1976) fundamentou uma estratégia de modelagem iterativa, dividindo a tarefa em três etapas: especificação do método a ser utilizado, estimativa dos parâmetros do modelo para realização das previsões e checagem do diagnóstico encontrado, visando avaliar a robustez do modelo. O processo é repetido até que o melhor método seja encontrado. A Figura 8 ilustra os procedimentos Box-Jenkins.



Fonte: Elaborada pelo autor (2020) a partir de Box & Jenkins (1976).

Segundo Werner & Ribeiro (2003), os modelos Box-Jenkins, são genericamente conhecidos por ARIMA (*Auto Regressive Integrated Moving Average*) e na literatura em português por Autorregressivos Integrados de Médias Móveis, sendo composta por modelos matemáticos que captam a correlação em séries temporais entre seus valores, e com base nesse comportamento, realiza previsões futuras. Os modelos ARIMA resultam da combinação de três componentes denominados “filtros”: o componente autorregressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA, do inglês *Moving Average*).

Para Hoshmand (2010) existem diversas razões para utilizar o método Box-Jenkins para a realização de previsões: por meio de um algoritmo matemático complexo, o método é capaz de entender uma grande variedade de padrões de dados, uma tarefa que exigiria muito se caso fosse feita individualmente, dado um conjunto de dados, o melhor modelo pode ser encontrado, o que economizaria tempo e recursos para a empresa. Essa abordagem de previsão é capaz de lidar com padrões

de dados complexos usando regras bem especificadas, provendo previsões confiáveis que podem ser testadas por meio de modelos estatísticos.

De acordo com Chase (2013), a abordagem Box-Jenkins incorpora elementos chave presentes nas técnicas de séries temporais e de regressões. Como resultado, os analistas que adotam essa técnica, devem ter um bom conhecimento em modelos de regressão antes de aplicar a abordagem Box-Jenkins.

Mesmo a notação teórica sendo sofisticada e complexa, aplicar os modelos ARIMA para fazer previsões não é tão difícil, especialmente depois dos avanços tecnológicos que permitiram automatizar os procedimentos Box-Jenkins usando pacotes de software e linguagens de programação (CHASE, 2013).

O modelo Box-Jenkins tem como principal assunção que o padrão de qualquer série temporal pode ser enquadrado em um dos três modelos descritos abaixo (BOX & JENKINS, 1976):

1. Auto regressivo (AR): Este modelo descreve o valor por meio da observação de uma série temporal, distribuindo pesos em observações anteriores. Onde p , que se refere ao peso, denota a parte auto regressiva.
2. Média Móvel (MA): Semelhante ao modelo Auto Regressivo (AR), este modelo define a observação por meio da atribuição de peso aos ruídos de observações anteriores. Onde q denota a parte da média móvel. O processo de MA relaciona o valor atual com o erro prévio do valor q .
3. Modelo híbrido ou modelo ARMA: É uma combinação dos dois modelos anteriores. Considera os valores passados e os ruídos do presente e do passado, onde d denota o grau primário de diferenciação envolvido. O modelo é formado utilizando uma combinação apropriada auto regressiva e de médias móveis. Sua aplicação limita-se a séries estacionárias, em que não haja variação de tendência e que as médias e variância sejam constantes. E, é importante observar que séries temporais reais podem não seguir um padrão estacionário.

Os modelos descritos acima podem ser utilizados quando uma série é estacionária, ou seja, quando suas características estatísticas, como média,

covariância e variância permanecem constantes ao longo do tempo (HANKE & WICHERN, 2014).

Existem diversas técnicas disponíveis para determinar se uma série é estacionária. Essas técnicas também são conhecidas como teste de raiz unitária, sendo o modelo mais utilizado o teste Dickey-Fuller. Um detalhamento maior desta técnica pode ser encontrado no trabalho de Dickey, Bell & Miller (1986).

Quando uma série é descrita como não estacionária, uma transformação é realizada por meio do processo de diferenciação dos dados com o objetivo de torná-la estacionária (MAKRIDAKIS, WHEELWRIGHT & HYNDMAN, 1997). Para tanto, complementando os modelos anteriormente descritos, o componente de integração (I), denotado por d é acrescido ao modelo, resultando no modelo ARIMA.

Os modelos ARIMA (Box & Jenkins, 1976) são geralmente representados pelos parâmetros p , d e q . O primeiro diz respeito a quantidade de termos AR, o segundo a quantidade de diferenciações necessárias para transformação de uma série não estacionária em estacionária, e por fim o último, descreve o número de médias móveis utilizadas no modelo. Este modelo pode ser representado pela Equação 9 descrita por Shukur & Lee (2015, p. 639):

$$\begin{aligned}
 & \left\{ \begin{array}{c} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \\ \mathbf{AR}(p) \end{array} \right\} \\
 & \times \left\{ \begin{array}{c} (1 - \Phi_1 B^{2s} - \Phi_2 B^2 - \dots - \Phi_p B^{Ps}) \\ \mathbf{AR}_s(P) \end{array} \right\} \left\{ (1 - B)^d \right\} \left\{ (1 - B^S)^d Y_t \right\} \\
 & \left\{ \begin{array}{c} I(d) \\ \mathbf{I}_s(D) \end{array} \right\} \\
 & = \left\{ \begin{array}{c} (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \\ \mathbf{MA}(q) \end{array} \right\} \\
 & \times \left\{ \begin{array}{c} (1 - \Theta_1 B - \Theta_2 B^{2s} - \dots - \Theta_q B^{Qs}) \\ \mathbf{MA}_s(Q) \end{array} \right\} \alpha_t
 \end{aligned} \tag{9}$$

Onde:

- $AR(p)$ é a ordem p do componente auto regressivo;
- $MA(q)$ é a ordem q do componente de médias móveis
- $I(d)$ é a ordem d da diferenciação não sazonal;
- $AR_s(P)$ é a ordem P do componente auto regressivo sazonal;

- $MA_s(Q)$ é a ordem Q do componente de médias móveis sazonal;
- $I_s(D)$ é diferenciação sazonal de ordem D e s é o período do padrão sazonal;
- B^i é a ordem i do operador de retrocesso; e
- $\phi, \Phi, \theta, \Theta$ são os parâmetros do modelo ARIMA.

Nos modelos ARIMA, coeficientes de auto correlação identificam a associação de uma variável em um período com a mesma variável em outro período, como por exemplo, a correlação da variável com ela mesma. O coeficiente de correlação pode ser encontrado por meio da Equação 10 descrita por Makridakis, Wheelwright & Hydman (1997, p. 39):

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (10)$$

Onde:

- r_k = coeficiente de correlação da observação Y_t com a observação Y_{t-k} ;
- k = representa a ordem do coeficiente de correlação;
- n = ordem da observação;
- Y_t = demanda no período observado (t); e
- \bar{Y} = média das observações da série temporal.

Segundo Kolassa & Siemesen (2016), estes modelos podem ser vistos como uma generalização dos modelos baseados em suavização exponencial. Métodos como a suavização exponencial são úteis para previsões de curto prazo, mas são ineficazes quando os padrões de dados estudados são mais complexos e combinam fatores como tendência, sazonalidade, flutuações cíclicas e aleatórias. Para lidar com esses cenários, é necessário adotar um método mais compreensível e que não tenha um grau de complexidade muito elevado. O método de regressão, apesar de sua facilidade de interpretação por conta das representações gráficas, torna este também seu principal ponto fraco, pois depende de visualizações e análises estatísticas para determinar o melhor modelo para a predição, dependendo de seu objetivo. Podemos concluir que o processo iterativo de determinação do melhor modelo a ser utilizado é custoso sob a ótica financeira e consome muito tempo.

Na Seção 3.3 serão detalhas técnicas de Aprendizagem de Máquina que podem ser utilizadas para a realização de previsões.

3.3. TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

Para Mitchell (1997), um algoritmo de aprendizagem é um conjunto de instruções que são necessárias para um computador realizar um aprendizado a partir de dados de entrada. Ainda, segundo o autor, um programa de computador aprende a partir de sua experiência com suas tarefas. Seu desempenho é avaliado por meio de parâmetros e o ciclo é reiniciado até que os erros encontrem seus valores mínimos.

Na Seção 3.2 observamos que os modelos matemáticos que conduzem as técnicas quantitativas são aplicados após uma análise exploratória realizada no conjunto de dados disponíveis, e o responsável por realizar esta atividade acaba avaliando a eficiência e eficácia do modelo baseado em algumas assunções. Em muitos casos, um modelo único é considerado, resultado de uma análise comparando os diversos modelos disponíveis e baseando-se na interpretação - da pessoa que realiza a atividade - do contexto. Segundo Breiman (2001), esse processo está relacionado à cultura de modelagem de dados. Para estudos focados em atividades de predição, novas tecnologias se popularizaram, entre elas o Aprendizado de Máquina (do inglês *Machine Learning*), tendo um desempenho relevante tanto na análise de dados atuais quanto para determinação do modelo mais apropriado para transformar esses dados em previsão (CLARK, 2013). Para Breiman (2001), essa perspectiva está mais relacionada à cultura de modelagem algorítmica.

Para Lantz (2015), os algoritmos de aprendizagem de máquina são divididos em categorias de acordo com seu propósito. Um modelo preditivo é usado para atividades que envolvem prever um valor usando outros valores presentes no conjunto de dados. O algoritmo, nesse caso, tenta descobrir e modelar o relacionamento entre o elemento alvo (o valor a ser previsto) e os preditores. O processo de treinamento de um modelo preditivo é chamado de aprendizado supervisionado. Ainda segundo o autor, a supervisão não está relacionada com a necessidade de um envolvimento

humano, mas sim no fato de que os valores alvo apresentam um caminho para que a máquina saiba o quão bem ela aprendeu a tarefa que lhe foi repassada.

Uma das tarefas de previsão relacionada ao aprendizado supervisionado é a classificação. Neste caso, o elemento alvo pode ser previsto de maneira categórica por meio de sua classe, que é então dividida entre níveis. Uma classe pode ter dois ou mais níveis e os níveis podem ou não ser ordinários. Um exemplo de tarefa de classificação, seria prever se um determinado e-mail se trata de uma mensagem publicitária, ensinando a máquina a reconhecer o conteúdo do e-mail.

Outra tarefa do aprendizado supervisionado³, trata-se da predição, ou regressão. Ao realizar previsões numéricas, é comum a utilização de modelos lineares que comportem os dados de entrada, utilizados para treinar o modelo. Esses modelos são amplamente utilizados em previsões pois torna-se possível quantificar em termos exatos a associação entre os valores de entrada (preditores) e o alvo, incluindo a magnitude e a incerteza sobre o relacionamento (LANTZ, 2015).

Nas próximas seções serão detalhadas algumas técnicas de aprendizagem de máquina aplicáveis no contexto da pesquisa.

3.3.1. Redes Neurais Artificiais

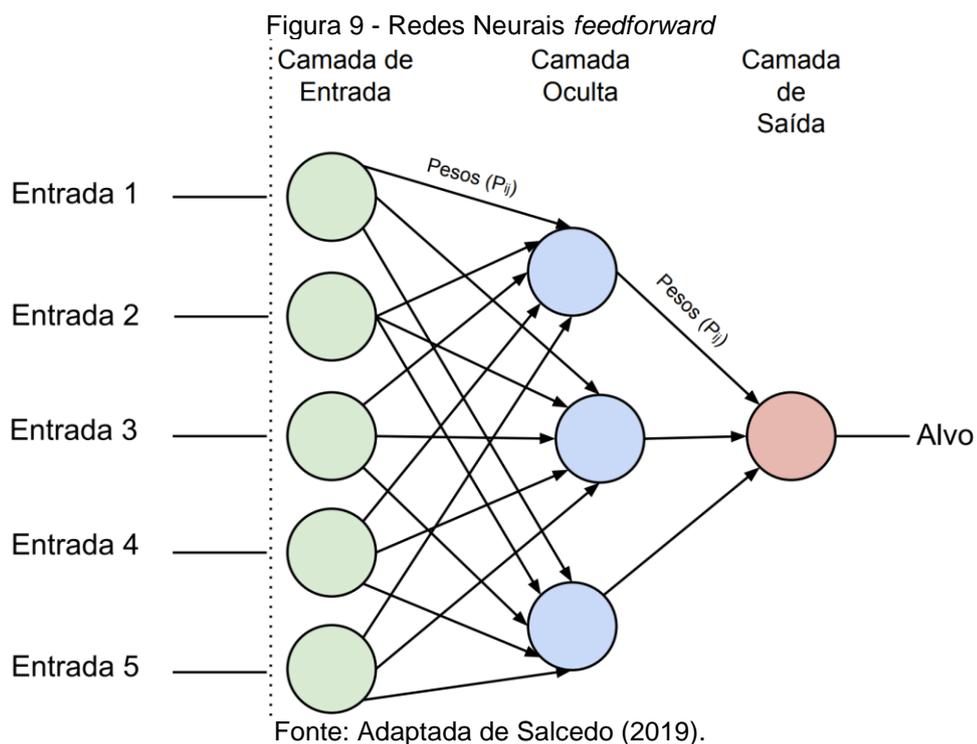
As Redes Neurais Artificiais (RNA) foram inicialmente desenvolvidas como uma tentativa de estudar e compreender como o cérebro humano funciona e originalmente eram utilizadas em disciplinas como neurociência e linguística

Uma RNA pode ser descrita como uma rede de “neurônios” que é organizada em camadas. Os preditores (ou variáveis de entrada) formam a base inferior e as previsões (saídas) criam a camada superior. É possível que também sejam formadas por camadas intermediárias contendo “neurônios ocultos”. Uma vez que essa camada

³ Existe também o aprendizado não-supervisionado, também chamado de aprendizado estatístico. Os conceitos sobre estes temas não serão detalhados no presente estudo, por não serem utilizados para as tarefas de predição. Mais detalhes podem ser encontrados no trabalho de Gutierrez (2015).

esteja presente, a RNA se torna então não-linear (HYNDMAN & ATHANASOPOULOS, 2018).

Em contraste com a linearidade dos modelos baseados em ARIMA para previsões, que são dependentes de suposições estatísticas que acabam por diminuir sua capacidade de prever um comportamento que não foi previamente registrado, as RNA tem como objetivo simular o comportamento de células neurais naturais, tornando-se orientadas a dados e auto adaptáveis de acordo com o processo de aprendizado na qual são submetidas e não necessitam de conhecimento prévio dos relacionamentos entre os métodos de previsão e os dados disponíveis (CRODA, ROMERO & MORALES, 2018). Com o avanço da tecnologia e a popularização do uso e computadores em ambientes corporativos, as RNA se tornaram alternativas viáveis para resolver uma grande variedade de problemas (LIN & WU, 2010). Um exemplo de RNA é exibido na Figura 9.



A Figura 9 ilustra uma rede neural *feedforward* (sem realimentação) também conhecida como MLP (do inglês *Multilayer Perceptron*), contendo cinco entradas, uma camada oculta com três neurônios ocultos. Para Ramasubramanian & Singh (2019), as RNA possuem três componentes principais:

- Arquitetura: Configuração do número de camadas, matriz de pesos, viés e conexões;
- Regras: Mecanismo que regula como os neurônios se comportam com relação as respostas de um para o outro;
- Regra de aprendizado: A forma em como os pesos atribuídos as redes neurais se alteram de acordo com o tempo.

No exemplo de RNA da Figura 9, o sinal de entrada percorre a rede em sentido único e os neurônios da mesma camada não são conectados. Cada camada de nós recebe informações de entrada das camadas anteriores. As saídas dos nós alimentam os nós seguintes. As entradas para cada nó são combinadas usando uma combinação linear ponderada. A Equação 11 descrita por Borovkyh, Bohte & Oosterlee (2018, p. 4):

$$a^1(i) = \sum_{j=1}^t \omega^1(i,j)x(j) + b^1(i) \quad (11)$$

Onde:

- x representa os dados de entrada;
- $\omega^1 \in \mathbb{R}^{M_1 \times t}$ representa as ponderações;
- $b^1 \in \mathbb{R}^{M_1}$ representa o viés;
- $a^1(i)$ representa a saída da rede.

O resultado então é modificado por uma função não linear antes de se transformar em saída, conforme a Equação 12 descrita por Hyndman & Athanasopoulos (2018, p. 444), que apresenta uma função sigmoide.

$$s(z) = \frac{1}{1+e^{-z}}, \quad (12)$$

A Equação 11 trata-se de uma função de ativação, responsável por processar o sinal de saída da função linear das entradas dos dados e seus respectivos pesos. A função sigmoide, exibida na Equação 12, pode ser definida como uma função crescente, responsável por balancear o comportamento linear e não linear dos dados, onde $s(z)$ representa o parâmetro de inclinação da função.

Existem outros tipos de arquiteturas de RNA utilizadas para tarefas de previsão, a exemplo das já citadas por Zhang, Patuwo & Hu (1998): Redes Neurais Recorrentes

de Hopfield (1982), Mapas de Kohonen (1982), Redes Neurais com Funções de Ativação de Base Radial (Chng *et al.*, 1996), *Ridge polynomial networks* (Shin & Ghosh, 1995) e *Wavelet Networks* (Zhang & Benveniste, 1992).

De acordo com Lantz (2015), as Redes Neurais Recorrentes têm sido utilizadas para uma variedade muito grande de aplicações e tem se tornado muito populares. Porém, segundo o autor, para desenvolver e aplicar esse tipo de RNA, geralmente, são necessários recursos computacionais especializados ou acesso a servidores *cloud*. Por outro lado, RNA mais simples como a *feedforward* são capazes de lidar com muitas tarefas do mundo real e requerem menos recursos computacionais, e por essa razão, e por estar implementada no R, essa arquitetura foi escolhida para esta pesquisa.

Algumas facilidades implementadas em linguagens de programação e softwares, permitem que, por meio da avaliação do conjunto de dados e suas variáveis, um modelo mais adequado seja sugerido para utilização.

A seguir, será apresentada a técnica de Aprendizagem de Máquina baseada em Modelos Lineares Generalizados.

3.3.2. Modelos Lineares Generalizados

Em estudos de natureza experimental ou observacional, encontramos problemas em que o principal objetivo torna-se entender a relação entre variáveis e, analisar a influência que uma ou mais variáveis (explicativas), tem sobre uma variável de interesse. Em estatística, o problema é abordado por meio de um estudo de um modelo de regressão, que seja capaz de relacionar a variável de interesse com as variáveis classificadas como explicativas (TURKMAN & SILVA, 2000).

O estudo proposto por Nelder & Wedderburn (1972) estabelece que modelos lineares podem ser generalizados para outros padrões por meio de uma função de ligação, que especifica de forma mais complexa o relacionamento entre duas variáveis. Esse método permite que a regressão seja aplicada para praticamente qualquer tipo de dado. Essa generalização é capaz de produzir uma previsão linear (LANTZ, 2015).

Segundo Paula (2013), a ideia principal seria permitir a abertura de um leque de opções para a distribuição da variável resposta, permitindo que pertença a família exponencial de distribuição, permitindo uma maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear.

Os Modelos Lineares Generalizados (MLG) estendem a ideia de uma regressão linear ordinária para outros tipos de distribuição de variáveis de resposta em uma família exponencial, assumindo que a variável dependente é gerada por uma distribuição de família exponencial. A título de conhecimento, as famílias exponenciais incluem normal, binomial, Poisson e distribuições *gamma*, entre outras, que não serão detalhadas no presente estudo. Um MLG é descrito por Ramasubramanian & Singh (2019, p. 330) por meio da Equação 13:

$$E(Y) = \mu = g^{-1}(X\beta) \quad (13)$$

Onde:

- $E(Y)$ é o valor esperado de Y ;
- $X\beta$ é o preditor linear, uma combinação de parâmetros desconhecidos β ;
- g é a função de ligação.

Os modelos de regressão generalizados têm forte influência estatística e são um dos primeiros algoritmos de aprendizagem de máquina por sua simplicidade e eficiência. Com o passar dos anos, novas técnicas de regressão surgiram, o que acabaram se desdobrando em abordagens e aplicações mais complexas (RAMASUBRAMANIAN & SINGH, 2019). Na próxima seção, será apresentado outro algoritmo baseado em Florestas Aleatórias.

3.3.3. Florestas Aleatórias

As Florestas Aleatórias são combinações de árvores de preditores de forma que cada árvore depende de valores que tem origem em um vetor de amostras aleatórias construído de forma independente e com a mesma distribuição para todas as árvores da floresta. A generalização de erros para as florestas converge para o

limite de árvores enquanto a floresta cresce. Essa generalização de erros dos classificadores das árvores dependerá da influência das árvores (individualmente falando) presentes na floresta e a correlação entre elas (BREIMAN, 2001).

Em outros métodos de árvore de decisão, cada nó é dividido usando o melhor nó entre todas as variáveis. Em uma Floresta Aleatória, cada nó é separado usando o melhor entre um conjunto menor de preditores selecionados aleatoriamente em um determinado nó (LIAW & WIENER, 2002). Essa estratégia contra intuitiva demonstra ter um desempenho melhor se comparada com outros classificadores, incluindo análise discriminatória, máquina de suporte de vetores e redes neurais, por sua robustez (BREIMAN, 2001).

O algoritmo das florestas aleatórias, é descrito por Liaw & Wiener (2002, p. 18):

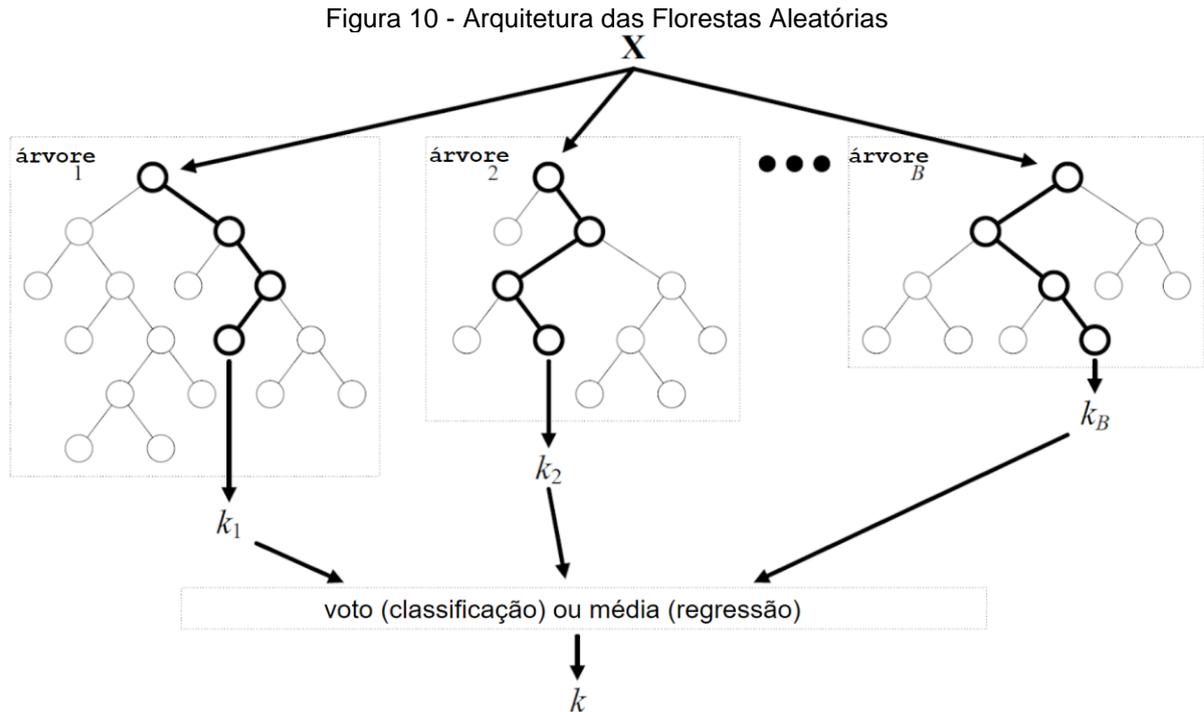
1. Desenhar $\eta_{árvores}$ de amostras *bootstrap*⁴ dos dados originais;
2. Para cada amostra *bootstrap*, criar uma árvore “não-podada” de regressão ou classificação, com a seguinte modificação:
 - a. Para cada nó, ao invés de escolher a melhor divisão entre todos os preditores, coletar uma amostra aleatória m_{try} dos preditores e escolher a melhor separação entre aquelas variáveis.
3. Prever os dados agregando as predições das árvores $\eta_{árvores}$ (maioria dos votos para classificação, média para regressão).

Já Breiman (2001, p. 6) define as Florestas Aleatórias como:

“[...] um classificador que consiste em uma coleção de classificadores estruturados em árvore $\{h(\mathbf{x}, \theta_\kappa), \kappa = 1, \dots\}$ onde os $\{\theta_\kappa\}$ são vetores aleatórios distribuídos de forma independente e cada árvore emite um voto unitário para a classe mais popular na entrada \mathbf{x} ”

Essa estrutura pode ser melhor entendida a partir da Figura 10, que demonstra a arquitetura das florestas aleatórias.

⁴ Definida por Alves (2013, p. 11) como “métodos computacionais intensivos que usam reamostragem para calcular as medidas de incerteza dos estimadores [...]”.



Na Figura 10, observamos X como um vetor de características representando a amostra a ser classificadas e então, temos a entrada de dados como origem das construções das árvores aleatórias, onde cada ramificação destaca o caminho escolhido para o algoritmo para se chegar no resultado esperado. Ao final do percurso, caso se trate de uma tarefa de classificação, teremos a avaliação da maioria dos votos para aquele resultado, e no caso de regressão, teremos a média encontrada ao final do percurso de cada árvore.

A aplicação desse modelo se mostra bem sucedida em uma grande variedade de áreas, incluindo tarefas de predição (VERIKAS *et al.*, 2016). Por esse motivo, torna-se uma técnica promissora para o objetivo desta pesquisa, levando em consideração a disponibilidade de dados e ao resultado esperado.

Este estudo vai ao encontro do entendimento de Hoshmand (2010), ao assumir que o processo de modelagem iterativa proposto por Box-Jenkins é o mais apropriado para realização de predição, pois sua metodologia permite perceber uma grande variação de padrões de dados, além da possibilidade de escolher o melhor modelo a ser empregado, de acordo com o conjunto de dados disponíveis.

Na próxima seção, serão apresentadas as métricas para medição de acurácia de predições. O entendimento destas métricas é de vital importância para avaliação da eficiência do modelo, e irão nortear o processo de determinação do melhor modelo.

3.4. MÉTRICAS PARA MEDIR ACURÁCIA DE PREDIÇÕES

Para que seja possível mensurar a precisão das técnicas de predição de demandas, torna-se necessária a utilização de métricas que forneçam a comparação entre os dados previstos por um determinado método e os dados observados em um determinado horizonte temporal. Analisando os resultados gerados pelos cálculos de erros de previsão, é possível avaliar a qualidade do ajuste do modelo testado.

A seguir, serão detalhas as métricas de Erro Médio Absoluto, Raiz Quadrada do Erro Médio, Erro Percentual Absoluto Médio e o Coeficiente de Determinação.

3.4.1. Erro Médio Absoluto

Uma das métricas mais básicas utilizadas para avaliar um modelo é o Erro Médio Absoluto (EMA) também conhecida por MAE, do inglês *Mean Absolute Error* (RAMASUBRAMANIAN & SIGH, 2019). O MAE tem como principal objetivo avaliar, em média, o quão distante a predição está do valor medido (LANTZ, 2015). A Equação 14 demonstra a fórmula utilizada para o seu cálculo, que considera a média do valor absoluto dos erros, uma vez que os erros são apenas a diferença entre os valores da previsão e os valores reais observados:

$$MAE = \left(\frac{1}{n}\right) \sum_{t=1}^n |y_t - x_t| \quad (14)$$

Onde:

- y representa o valor observado no momento t ;
- x representa o valor previsto para o momento t ;

- n representa os valores observados desde $t = 1$.

3.4.2. Raiz Quadrada do Erro Médio

A Raiz Quadrada do Erro Médio (RQEM) ou *Root Mean Square Error* (RMSE) é uma das métricas mais populares no que diz respeito a avaliação de erros contínuos em modelos (RAMASUBRAMANIAN & SING, 2019). Sua fórmula é descrita na Equação 15:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{t=1}^n (y_t - x_t)^2} \quad (15)$$

A representação dos elementos é semelhante à Equação 13, mas nesse caso cada erro absoluto é elevado ao quadrado antes de ser somado e a raiz quadrada do erro médio quadrático é extraída antes de ser retornada.

3.4.3. Erro Percentual Absoluto Médio

Já o Erro Percentual Absoluto Médio (EPAM) ou *Mean Absolute Percentage Error* (MAPE), tem como objetivo transformar o MAE em valor percentual, conforme a Equação 16:

$$MAPE = \left(\frac{1}{n}\right) \sum_{t=1}^n \left| \frac{y_t - x_t}{y_t} \right| \times 100 \quad (16)$$

3.4.4. Coeficiente de Determinação

Outra métrica bastante popular utilizada para mensurar a acurácia de modelos baseados em regressões lineares é o coeficiente de determinação, também conhecido como R^2 . Essa métrica descreve a proporção da variação de Y que pode ser explicada

por variações em X de acordo com o modelo de regressão utilizado. A correlação é denotada por um valor entre 0 e 1. O valor extremo e positivo, indica uma relação linear perfeita, enquanto uma correlação próxima de zero indica a inexistência de uma relação linear. A equação 17 demonstra como o valor é calculado.

$$R^2 = 1 - \frac{\sum_{t=1}^n (Y_t - x_t)^2}{\sum_{t=1}^n (\bar{Y}_t - x_t)^2} \quad (17)$$

Onde:

- $\sum_{t=1}^n (Y_t - x_t)^2$ é a Soma Total dos Quadrados (STQ);
- $\sum_{t=1}^n (\bar{Y}_t - x_t)^2$ é a Soma dos Quadrados Explicada (SQE);

Podemos entender o R^2 como uma medida da qualidade geral do modelo de regressão utilizado (LEWIS, 2017).

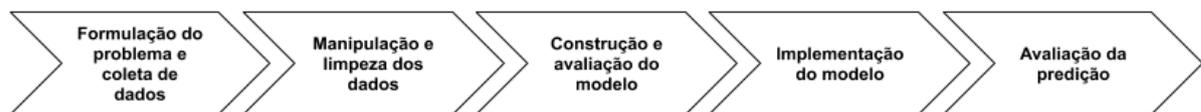
No Capítulo 4, os procedimentos metodológicos e o ambiente do estudo serão detalhados.

4. PROCEDIMENTOS METODOLÓGICOS

Quanto ao processo de coleta de dados, Lakatos & Marconi (1992) afirmam que podemos obter dados por meio de dois processos: documentação direta e indireta. O primeiro processo refere-se a coleta dos dados no local que os fenômenos se manifestam usando a pesquisa de campo ou a pesquisa de laboratório. A segunda utiliza dados coletados por outras pessoas, tendo como fonte um material já elaborado. Os dados deste trabalho foram obtidos por meio da documentação direta, ou seja, do acervo pertencente à administração da empresa objeto desta pesquisa, utilizando uma abordagem quantitativa, tendo como principal objetivo a utilização de dados históricos para a previsão de dados no futuro. Por fim, esta pesquisa assumirá o perfil de pesquisa exploratória permitindo que seus resultados sejam explorados em trabalhos futuros.

As etapas desta pesquisa, seguem os passos sugeridos por Hanke & Wichern (2014) e ilustrados na Figura 11.

Figura 11 - Processos de desenvolvimento da pesquisa



Fonte: Adaptada de Hanke & Wichern (2014).

A primeira etapa deste processo é a formulação do problema que é representado neste estudo pelo problema de pesquisa. A segunda etapa descrita no processo, é a manipulação e limpeza dos dados. Para tanto, é necessário realizar a extração e integração dos dados para construção de um formato padrão. É comum que os dados, mesmo que sejam provenientes da mesma fonte, apresentem falhas ou uma organização inadequada para a aplicação de algoritmos. Essas condições geram diversos problemas durante o processo de descoberta do conhecimento e podem ocasionar na falha da análise pretendida. Neste sentido, é de vital importância a realização de etapas de pré-processamento buscando corrigir inconsistências além da possibilidade de agregar valor aos dados para que os resultados sejam confiáveis e de qualidade. Os problemas mais comuns relacionados a qualidade de dados, de

acordo com (Silva, Peres & Boscaroli, 2016) são ausência de valores, dados com ruídos, valores inconsistentes e redundância nos dados.

Para dar suporte ao negócio e fornecer inteligência para apoio à decisão, as empresas têm investido cada vez mais em ferramentas como o R para construção de algoritmos de aprendizagem de máquina e realizar análise dos conjuntos de dados disponíveis (FOWDUR *et al.*, 2018). A linguagem de programação R foi selecionada, por se tratar de uma suíte integrada de facilidades para manipulação de dados, realização de cálculos e visualização de gráficos. Trata dados de maneira eficiente facilitando seu armazenamento, pode realizar cálculos complexos por meio de listas ou matrizes, oferecendo uma coleção de ferramentas para análise de dados e para criação e visualização de gráficos.

A linguagem de programação R está disponível como um software livre e de código aberto, sob os termos GNU (*General Public License*). Seu código pode ser compilado e executado em grande variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS (R Core Team, 2015). A implementação foi realizada utilizando o software R *Studio* Versão 1.2.5042 (<https://www.r-project.org/>).

Para iniciar a segunda etapa do processo, foram extraídos da base de dados do ERP utilizado pela franquia, os relatórios referentes a vendas dos meses de setembro de 2019 a janeiro de 2020, período este que foi a janela autorizada pela franqueadora para abertura dos dados relativos a vendas. É importante ressaltar que a franqueadora optou por alterar o ERP utilizado nas unidades, por conta de decisões estratégicas relativas ao negócio e isso acabou por reduzir o horizonte de dados disponíveis para consulta, uma vez que, por razões técnicas, os dados do ERP anterior não puderam ser importados na base de dados do novo software.

Ademais, o período foi selecionado para contemplar as vendas do cardápio utilizado quando a pesquisa foi conduzida. Já os dados referentes as vendas nos meses de fevereiro de 2020 foram utilizados para comparar a margem de erro do modelo sugerido. A Figura 12, demonstra um recorte de um relatório gerencial das vendas diárias do mês de setembro de 2019.

Figura 12 - Relatório de movimentação de produtos por dia

| CÓDIGO | PRODUTO | DATA | QUANTIDADE |
|---------------|---|------------|------------|
| --- | Outras despesas, Descontos e Frete | | 0,00 |
| 1 | ICE LEMON - MATE ROYALE, POP MARACUJÁ | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - SICILIANO BLUE, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - CLASSIC PEACH, POP BLUEBERRY | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - BLACK MINT | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - FRUTAS VERMELHAS, POP MORANGO | 01/09/2019 | 4,00 Un |
| 1 | ICE LEMON - FRUTAS VERMELHAS, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - SUNSET, JEL ABACAXI | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - MENTA E LIMÃO, JEL ABACAXI | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - SUNSET | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - LYCHEE LUSH, POP LICHIA | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - MATE COM LIMÃO, POP BLUEBERRY | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - MENTA E LIMÃO, POP MORANGO | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - MATE COM LIMÃO, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - HIBISCO, POP MAÇA VERDE | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - RED HILL, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - MARACUCHÁ, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - GREEN LIGHT, POP MARACUJÁ | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - GREEN LIGHT, POP MAÇA VERDE, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - FRUTAS CÍTRICAS, POP BLUEBERRY | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - SICILIANO BLUE, POP MAÇA VERDE | 01/09/2019 | 2,00 Un |
| 1 | ICE LEMON - FRUTAS VERMELHAS, POP BLUEBERRY, POP LICHIA | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - MATE ROYALE, POP BLUEBERRY, POP LICHIA | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - FRUTAS VERMELHAS, POP MAÇA VERDE | 01/09/2019 | 1,00 Un |
| 1 | ICE LEMON - FRESH'N GINGER, POP BLUEBERRY | 01/09/2019 | 2,00 Un |
| TOTAL: | | | 3744,00 |
| TOTAL PÁGINA: | | | 34,00 |

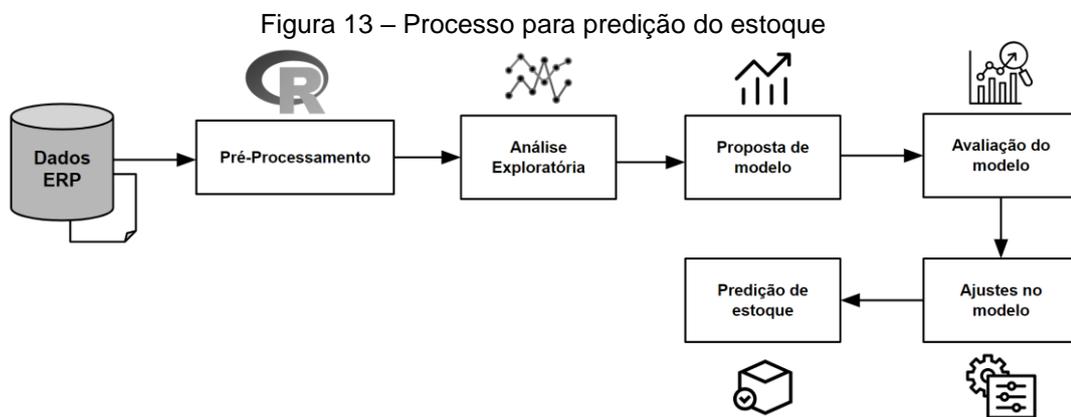
Fonte: ERP utilizado pela franquia (2020).

O software permite a exportação do arquivo para o formato .CSV (*Comma Separated Values*) e pode ser aberto por programas de edição de planilhas eletrônicas. O formato de exportação, conforme é possível observar na Figura 12, possui uma estrutura adequada para um relatório gerencial, porém, inadequada para aplicação para fins de predição. Segundo Silva, Peres & Boscarioli (2016), a importação de arquivos em R pode ser feita de duas formas: por meio da leitura do próprio arquivo ou a partir de bases de dados. O procedimento mais comum é a leitura de arquivos com valores separados por um identificador, como o CSV, que é um formato de exportação presente em ambientes de exploração de dados.

Focando no modelo preditivo proposto para vendas, um método de previsão foi selecionado de acordo com a disponibilidade de dados da unidade piloto pertencente

a rede de franquias de *bubble tea*. Os modelos de séries temporais e de regressão podem ser utilizados com eficiência para gerar previsões na maioria dos ambientes de negócio ou econômicos, porém, esses modelos nem sempre identificam os padrões dos dados exibidos em muitas séries temporais.

Portanto, a pesquisa empregou processo proposto por Box-Jenkins para seleção do modelo preditivo. Os tratamentos estatísticos do conjunto de dados – foram realizados utilizando a linguagem de programação R. A Figura 13 ilustra o processo realizado para predição do estoque.



Fonte: Baseada em Hanke & Wichern (2014).

Os dados dos produtos do cardápio disponíveis no ERP da empresa foram extraídos por meio de relatórios gerenciais das vendas diárias no período determinado para a pesquisa. Estes dados foram processados por uma implementação feita na linguagem de programação R, em que foram transformados para um formato adequado para posterior realização da análise exploratória, que teve com principal objetivo, identificar correlações entre variáveis internas e externas ao processo de vendas da unidade. Esta análise permitiu a criação de um novo conjunto de dados, contendo as variáveis de interesse identificadas durante a análise exploratória.

Após a criação do conjunto de dados, os modelos candidatos foram aplicados, produzindo um novo conjunto com os valores projetados para o período de um mês. Para avaliação dos modelos candidatos, foram verificados os erros de previsão, utilizando o MAE (Erro Médio Absoluto), MAPE (Erro Percentual Absoluto Médio), o RMSE (Raiz Quadrada do Erro Médio) e quando possível, o coeficiente de determinação (R^2). Para a adequação do modelo, foram realizados ajustes de parâmetros baseados nas avaliações e uma projeção do nível de estoque foi

calculada, considerando a quantidade necessária de cada insumo por meio de sua Ficha Técnica de Preparo (FTP) definida por Vasconcellos (2002, p. 36) como:

“Uma receita padronizada, na qual são discriminados todos os ingredientes utilizados em cada uma das preparações. É, portanto, um instrumento gerencial de apoio operacional, pelo qual se fazem o levantamento dos custos [...]”

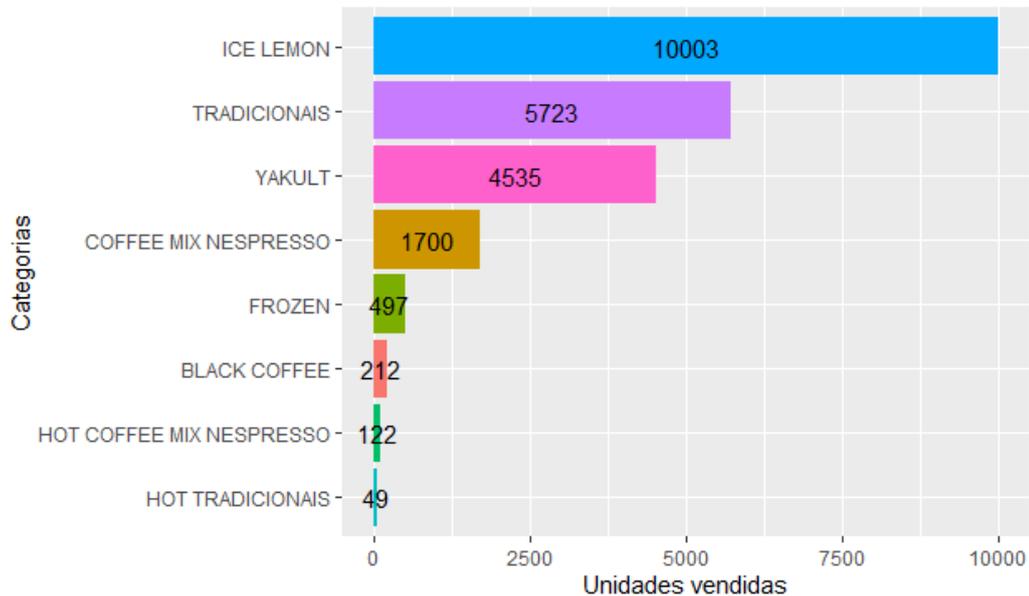
Por fim, utilizando o modelo proposto de predição de vendas, seu resultado foi comparado utilizando os dados das vendas do mês de fevereiro de 2020, com o objetivo de avaliar a eficiência e acurácia do modelo. Esse período foi selecionado, foi o horizonte determinado pela empresa, levando em consideração a autorização concedida para realização desta pesquisa, presente como Anexo deste documento.

4.1. Análise Experimental

O principal objetivo da análise das séries temporais neste estudo é estimar os itens que compõem a demanda (sazonalidade, ciclo, tendência e erro) com o objetivo de encontrar uma equação matemática que represente seu comportamento, como descrito em (CHOPRA; MENDL, 2016). Neste estudo, foram analisadas as vendas mensais durante os meses de setembro, outubro, novembro e dezembro de 2019 e janeiro e fevereiro de 2020.

A Figura 14 demonstra o total de vendas por categoria durante o período analisado.

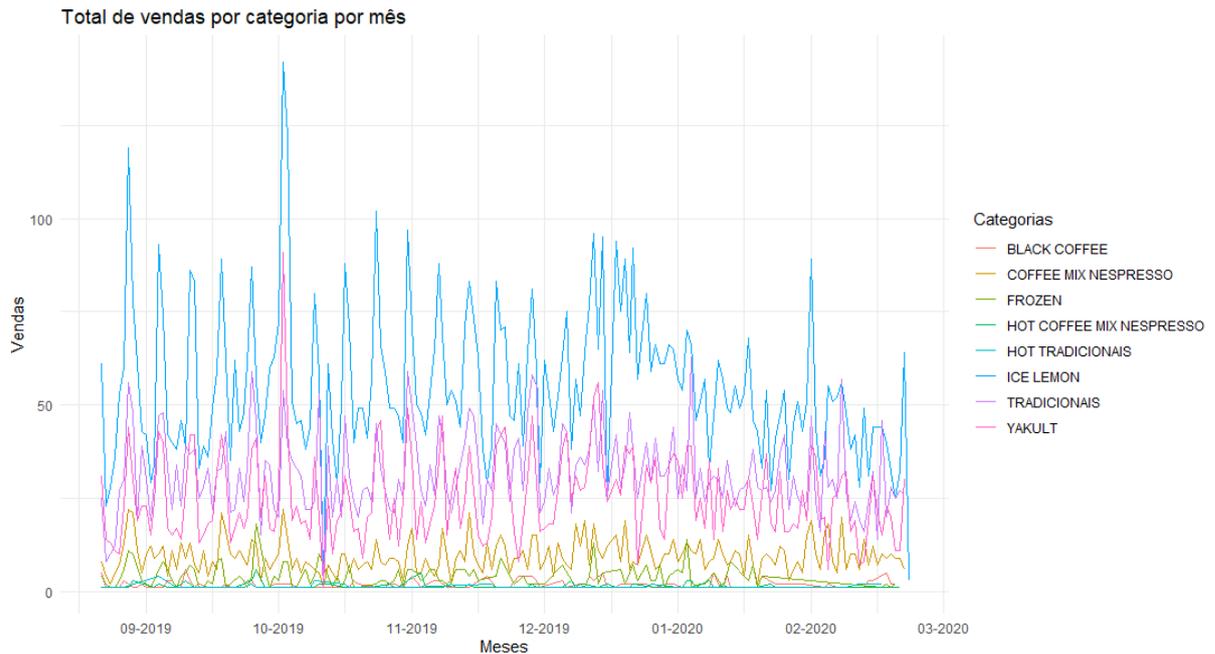
Figura 14 - Vendas de *bubble tea* por categoria
Total de vendas por categoria



Fonte: Elaborada pelo autor (2020) a partir dos dados da pesquisa.

O gráfico da Figura 14 apresenta as vendas totais durante o período observado, exibindo os grupos de produtos comercializados: *Black Coffee* (café espresso), *Coffee Mix Nespresso* (*bubble tea* a base de café), *Frozen* (espécie de *milkshake*), *Hot Coffee Mix Nespresso* (*bubble tea* a base de café em sua versão quente), *Hot Tradicionais* (*bubble tea* cremoso em sua versão quente), *Ice Lemon* (*bubble tea* refrescante a base de limão), *Tradicionais* (*bubble tea* cremoso) e *Yakult* (*bubble tea* a base de chá verde com leite fermentado). A Figura 15 exhibe o comportamento mensal das vendas por categoria no período observado e demonstra uma irregularidade, não sendo possível determinar se existe um padrão de sazonalidade. Este padrão poderia ser mais evidente se a massa de dados apresentasse ao menos um ou mais anos de vendas acumuladas, mas por conta das razões já exploradas anteriormente que envolvem troca do ERP e o formato dos dados, uma análise cíclica não foi possível.

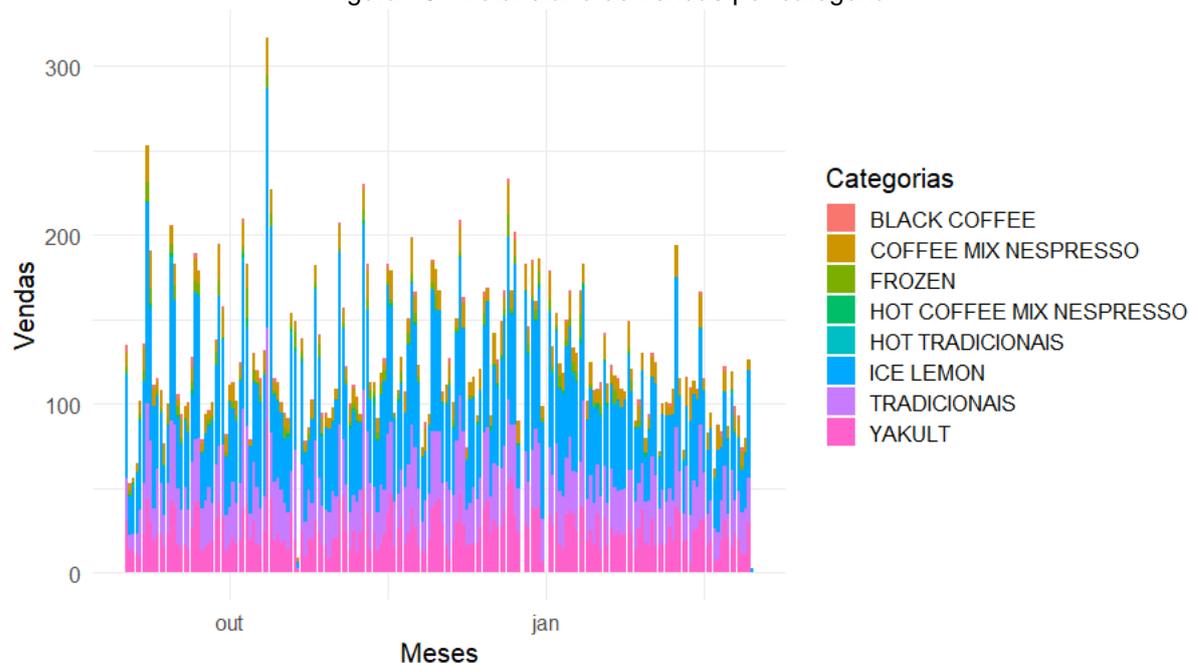
Figura 15 - Vendas mensais por categoria



Fonte: Elaborada pelo autor (2020).

Na Figura 16, podemos observar as vendas diárias por categoria, durante o período observado. A partir dessa ilustração é possível determinar visualmente as categorias de maior interesse do público consumidor, para que posteriormente essa visão seja comparada com os resultados da previsão.

Figura 16 - Total diário de vendas por categoria



Fonte: Elaborada pelo autor (2020).

Autores como van der Vorst *et al.* (1998) e van Donselaar *et al.* (2006) afirmam que as variações na demanda podem ser causadas por questões internas (promoções, redução de preços, liquidações etc.) ou questões externas (feriados, férias escolares, clima etc.). Para identificar se a série possui correlação com outras variáveis não visíveis ao observar sua variação, é importante entender quais fatores externos podem influenciar o comportamento do consumidor. A pesquisa de Granger (1978) evidenciava que o clima e suas variações possuem um impacto considerável no desempenho econômico de diversos setores e pesquisas mais recentes comprovam estatisticamente essa influência (DELL, JONES & OLKEN, 2013). A pertinência do clima com a venda de bebidas não alcoólicas é comprovada em pesquisas específicas que investigam essa relação (DIVAKAR *et al.*, 2005; GUSTAVSEN, 2005; RAMANATHAN & MULDERYMANS, 2010; ŠTULEC, PETLJAK & NALETINA, 2019).

Buscando evidenciar esta influência investigada em estudos correlatos, foram coletados dados sobre o clima de Foz do Iguaçu – PR (cidade onde a unidade investigada está instalada) na plataforma do INMET – Instituto Nacional de Meteorologia⁵, dos meses de setembro, outubro, novembro e dezembro de 2019 e janeiro e fevereiro de 2020. A Figura 17 exibe o relatório parcial exibido pela consulta.

Figura 17 - Dados históricos sobre o clima de Foz do Iguaçu – PR

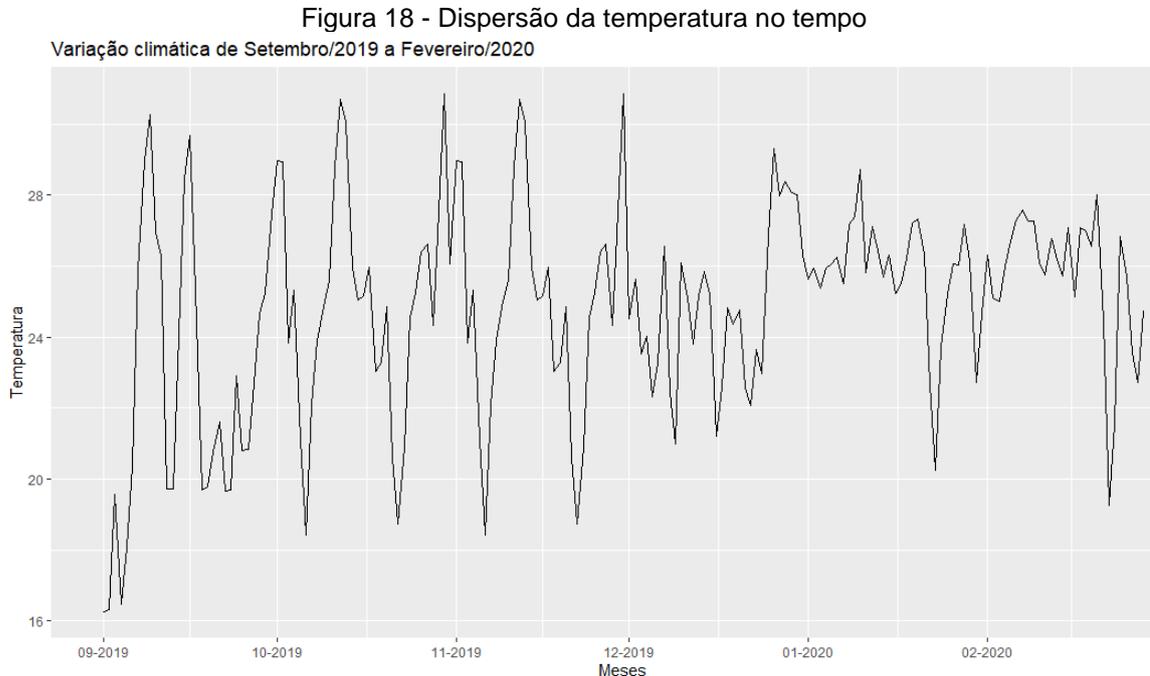
Data Inicial: 01/09/2019 Data Final: 29/02/2020

| Data | Hora | Temperatura (°C) | | | Umidade (%) | | | Pto. Orvalho (°C) | | | Pressão (hPa) | | | Vento (m/s) | | | Radiação | Chuva | |
|------------|------|------------------|-------|------|-------------|-------|------|-------------------|-------|------|---------------|-------|-------|-------------|------|----------|----------|----------------------|------|
| | | UTC | Inst. | Máx. | Mín. | Inst. | Máx. | Mín. | Inst. | Máx. | Mín. | Inst. | Máx. | Mín. | Vel. | Dir. (°) | Raj. | (kJ/m ²) | (mm) |
| 01/09/2019 | 00 | | 18.7 | 19.6 | 18.7 | 78 | 78 | 76 | 14.8 | 15.3 | 14.8 | 986.8 | 986.8 | 985.9 | 4.9 | 330 | 9.5 | -2.49 | 0.0 |
| 01/09/2019 | 01 | | 17.6 | 18.7 | 17.6 | 81 | 81 | 78 | 14.2 | 14.8 | 14.2 | 986.7 | 987.0 | 986.6 | 4.1 | 332 | 9.5 | -3.31 | 0.0 |
| 01/09/2019 | 02 | | 16.9 | | | 82 | | | 13.8 | | | 987.2 | | | | | | -0.06 | 0.0 |
| 01/09/2019 | 03 | | 16.0 | 16.9 | 16.0 | 84 | 84 | 82 | 13.2 | 13.8 | 13.2 | 987.4 | 987.5 | 987.1 | 2.3 | 336 | 5.2 | -3.52 | 0.0 |
| 01/09/2019 | 04 | | 15.5 | 16.0 | 15.5 | 85 | 85 | 84 | 12.9 | 13.2 | 12.9 | 987.2 | 987.5 | 987.1 | 2.2 | 341 | 4.7 | -3.12 | 0.0 |
| 01/09/2019 | 05 | | 14.8 | 15.5 | 14.8 | 85 | 85 | 85 | 12.4 | 12.9 | 12.3 | 987.2 | 987.4 | 987.1 | 2.0 | 351 | 3.8 | -3.49 | 0.0 |
| 01/09/2019 | 12 | | 15.4 | 15.7 | 14.4 | 86 | 88 | 86 | 13.0 | 13.4 | 12.3 | 989.1 | 989.3 | 988.0 | 3.1 | 321 | 7.4 | 634.0 | 0.0 |
| 01/09/2019 | 14 | | 18.4 | 19.6 | 17.2 | 75 | 82 | 75 | 13.9 | 15.6 | 13.8 | 989.5 | 989.6 | 989.0 | 5.1 | 340 | 9.3 | 2013. | 0.0 |
| 01/09/2019 | 15 | | 18.0 | 18.7 | 18.0 | 74 | 76 | 74 | 13.3 | 14.2 | 13.3 | 989.1 | 989.7 | 989.0 | 4.6 | 337 | 9.3 | 967.8 | 0.0 |
| 01/09/2019 | 16 | | 17.8 | 18.3 | 17.8 | 74 | 74 | 73 | 13.1 | 13.6 | 12.9 | 988.9 | 989.7 | 988.8 | 2.4 | 357 | 7.9 | 427.8 | 0.0 |
| 01/09/2019 | 20 | | 15.7 | 16.1 | 15.5 | 77 | 77 | 75 | 11.8 | 11.8 | 11.4 | 987.3 | 988.0 | 987.0 | 3.1 | 352 | 6.5 | 476.1 | 0.0 |
| 01/09/2019 | 21 | | 14.9 | 15.8 | 14.9 | 77 | 77 | 77 | 11.0 | 11.8 | 11.0 | 987.5 | 987.7 | 987.3 | 3.2 | 345 | 7.2 | 180.2 | 0.0 |
| 01/09/2019 | 22 | | 14.3 | 14.9 | 14.3 | 78 | 78 | 77 | 10.6 | 11.0 | 10.6 | 988.0 | 988.1 | 987.5 | 2.0 | 346 | 5.8 | 12.02 | 0.0 |
| 01/09/2019 | 23 | | 13.4 | 14.3 | 13.3 | 79 | 79 | 78 | 9.8 | 10.6 | 9.8 | 988.3 | 988.3 | 988.0 | 1.6 | 330 | 3.9 | -3.54 | 0.0 |

Fonte: INMENT (2020)

⁵ Consulta disponível em: http://www.inmet.gov.br/sonabra/pg_dspDadosCodigo_sim.php?QTg0Ng==

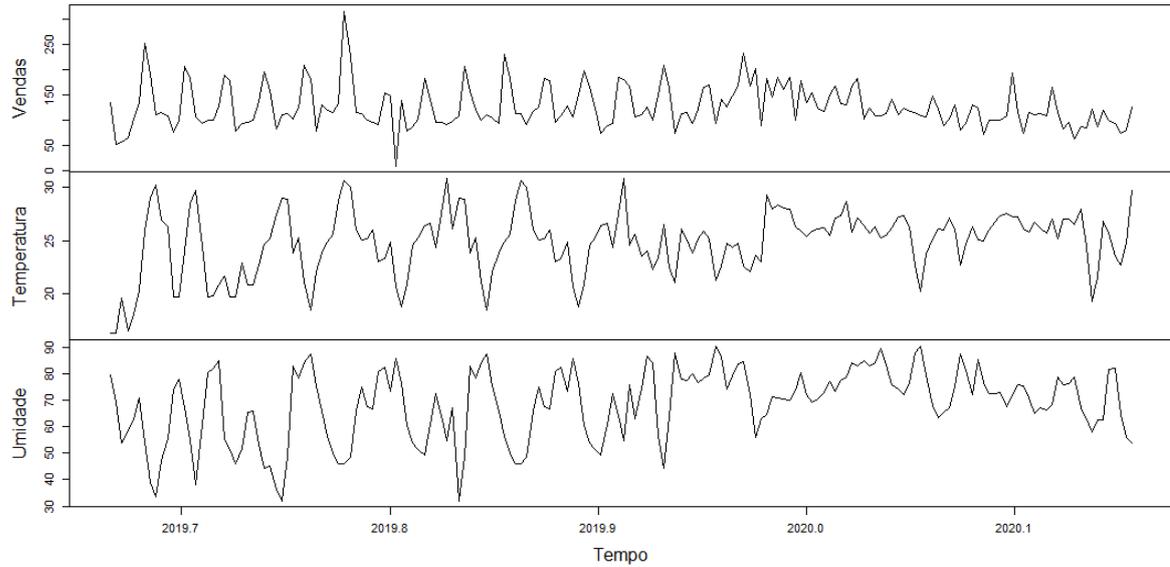
A possibilidade de download dos dados no formato CSV contribuiu para o pré-processamento e inclusão no conjunto de dados proposto para a análise. A Figura 18 ilustra a dispersão da temperatura no tempo.



Fonte: Elaborada pelo autor (2020).

De posse dessas informações, é possível distribuir essas informações em um gráfico, comparando a flutuação dos dados climáticos com a flutuação das vendas no período observado. As variáveis selecionadas para composição do conjunto de dados foram a temperatura instantânea e a umidade instantânea. A Figura 19 exibe as séries temporais dessas variáveis em conjunto com a série temporal de vendas, de todos os produtos do cardápio, com exceção de produtos sazonais (para datas comemorativas) ou edições limitadas que não fazem parte do cardápio de rotina da unidade.

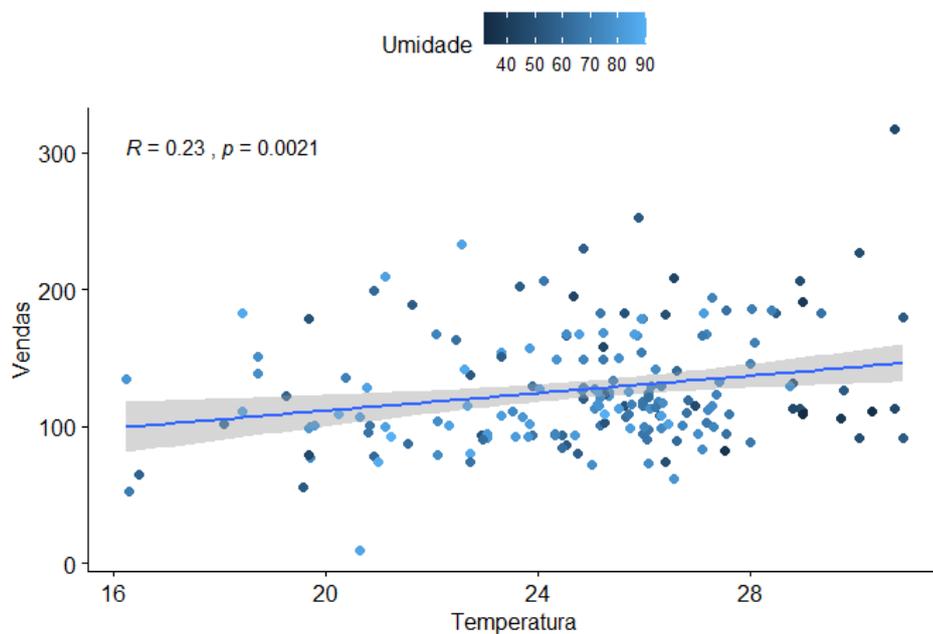
Figura 19 - Dispersão das vendas, temperatura e umidade no tempo



Fonte: Elaborada pelo autor (2020).

Apenas observando o gráfico na Figura 19 é possível afirmar que existe uma correlação de leve a moderada com relação a influência dos aspectos climáticos nas vendas. Para verificar a correlação de forma mais detalhada, foi calculado o coeficiente de relação de Pearson.

Figura 20 - Vendas de acordo com a temperatura



Fonte: Elaborada pelo autor (2020).

A hipótese mais provável ao visualizar o gráfico da Figura 20, é que a correlação seja leve, uma vez que os pontos relacionados as vendas estão próximas a linha central, quando a temperatura aumenta, mas não estão agrupados de forma uniforme e linear. Os tons de azul aplicados a umidade testam se esta variável se relaciona de alguma forma com a percepção da temperatura. A realização do cálculo do coeficiente de Pearson para compreender a correlação numérica entre essas variáveis (Vendas e Temperatura), demonstra que $R = 0.23$ e $p = 0.0021$. De acordo com Fernandes (1999), a interpretação é que:

- As correlações podem ser positivas ou negativas, tendo seus valores entre -1 e +1;
- A proximidade de +1 ou -1 representa uma forte correlação;
- Se os valores de X formam pares aleatoriamente com os valores de Y, a medida de correlação deve estar próxima de zero.

Na Figura 20, observa-se um coeficiente de correlação da amostra de Pearson denotado pelo valor de $R = 0.23$, logo, podemos afirmar que existe uma correlação positiva leve entre as duas variáveis. Uma hipótese nula descreve a inexistência da relação entre duas variáveis. Se o valor de p for igual a zero, significa que as variáveis são independentes e não possuem correlação. O valor de p encontrado, se mostra menor do que 0.05, então, é possível afirmar que existe uma significância estatística que fortalece a existência de uma hipótese alternativa investigada nesta pesquisa, de que a temperatura tem influência no número absoluto de vendas de *bubble tea*.

Ainda seguindo esta linha de raciocínio, temos uma outra hipótese, a de que a estação do ano pode influenciar a flutuação de vendas, já que verificamos uma relação positiva entre a temperatura e o número de bebidas comercializadas. Para que seja possível identificar essa relação, foi necessário a criação de uma nova coluna no conjunto de dados que represente a estação do ano de cada dia das vendas.

A Tabela 3 demonstra como essa estrutura retorna a informação e a Figura 21 ilustra o reflexo da informação em um recorte do conjunto de dados.

Tabela 3 - Estrutura dos dados de estações climáticas

| Estação climática | Referência |
|--------------------------|-------------------|
| Outono | 1 |
| Inverno | 2 |
| Primavera | 3 |
| Verão | 4 |

Fonte: Elaborada pelo autor (2020).

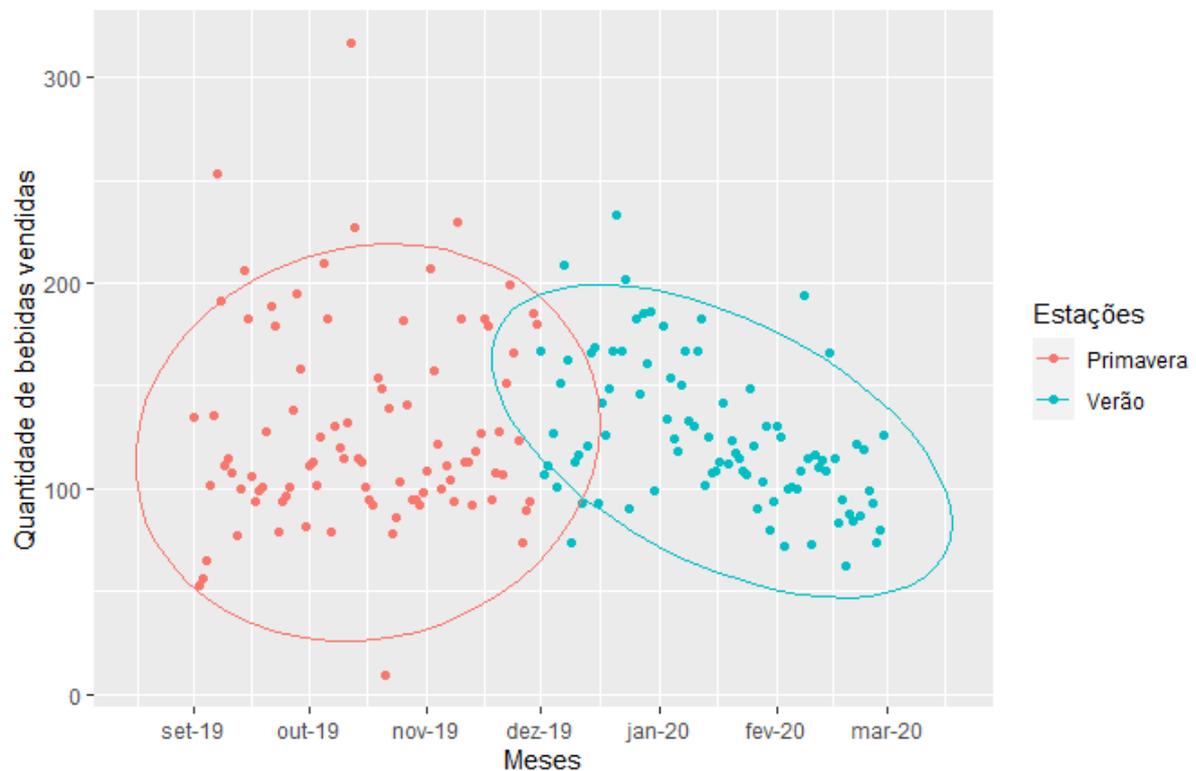
Figura 21 – Recorte do conjunto de dados contendo as estações

| | date | quantity | temp_inst | umid_inst | season |
|---|-------------|-----------------|------------------|------------------|---------------|
| 1 | 2019-09-01 | 135 | 16.24286 | 79.64286 | 3 |
| 2 | 2019-09-02 | 53 | 16.30000 | 69.30000 | 3 |
| 3 | 2019-09-03 | 56 | 19.56500 | 53.85000 | 3 |
| 4 | 2019-09-04 | 65 | 16.48095 | 58.23810 | 3 |
| 5 | 2019-09-05 | 102 | 18.07059 | 62.70588 | 3 |
| 6 | 2019-09-06 | 136 | 20.35714 | 70.71429 | 3 |
| 7 | 2019-09-07 | 253 | 25.88636 | 54.09091 | 3 |
| 8 | 2019-09-08 | 191 | 28.96667 | 39.08333 | 3 |
| 9 | 2019-09-09 | 111 | 30.26818 | 33.45455 | 3 |

Fonte: Elaborada pelo autor (2020).

Após a informação inserida, é possível criar um gráfico para visualizar a correlação entre as vendas e as estações do ano. A Figura 22 exibe o gráfico contendo essa informação. Neste caso, uma observação importante deve ser feita, por conta da limitação dos dados disponibilizados. Em um cenário ideal, seria necessário que o conjunto de dados refletisse um ano completo para que a correlação seja compreendida considerando o ciclo das estações e a flutuação das vendas durante todo o período. Neste estudo, o conjunto de dados criado não permite uma visão mais detalhada sobre os aspectos sazonais relacionados as estações climáticas. Logo, a Figura 22 evidencia apenas a correlação das vendas entre a primavera e o verão, que compreendem exatamente o período de 180 dias presente no conjunto de dados.

Figura 22 - Vendas de bebidas por estação do ano



Fonte: Elaborada pelo autor (2020).

Essa limitação se torna ainda mais evidente quando realizamos o cálculo do coeficiente de Pearson, retornando o $R = -0.033$ e o valor de $p = 0.6587$. Interpretando esses resultados, podemos verificar a existência de uma correlação negativa entre as variáveis, e o valor de p demonstra ser acima de 0.05, o que torna a hipótese nula. Nesse caso, é importante entender que os testes matemáticos apenas consideram a correlação entre essas duas estações na flutuação das vendas, o que torna o resultado inconclusivo por conta da ausência de dados de vendas das demais estações (outono e inverno).

Para entender melhor os padrões de consumo da unidade estudada, torna-se necessário levar em consideração os efeitos de calendário nas vendas dos produtos. Uma das hipóteses a ser explorada dentro destes efeitos, é a flutuação de vendas em fins de semana e feriados. Os dados coletados diretamente do sistema de gestão da unidade, não diferencia os efeitos de calendário, portanto, para que seja possível realizar esta análise, foi criada uma coluna no conjunto de dados a partir de uma lista contendo os feriados da cidade de Foz do Iguaçu – PR por meio do Decreto N° 26.945 de 8 de janeiro de 2019 que “Divulga os feriados e pontos facultativos para o exercício de 2019, para os órgãos e entidades da Administração Municipal Direta, Autárquica e

Fundacional, do Poder Executivo” (Foz do Iguaçu (PR), 2019) e do Decreto N° 27.806 de 6 de janeiro de 2020 que “Divulga os feriados e pontos facultativos para o exercício de 2020, para os órgãos e entidades da Administração Municipal Direta, Autárquica e Fundacional, do Poder Executivo” (Foz do Iguaçu (PR), 2020).

Essa coluna verifica se a data das vendas totais se encaixa nestes casos, retornando um valor *booleano* caso seja confirmado um feriado em uma determinada data. Um procedimento semelhante foi realizado para identificação dos finais de semana, sendo considerados no contexto do negócio: sexta-feira, sábado e domingo como dias pertencentes a esta categoria. A Figura 23 exhibe a estrutura de dados idealizada para que essa atividade seja executada:

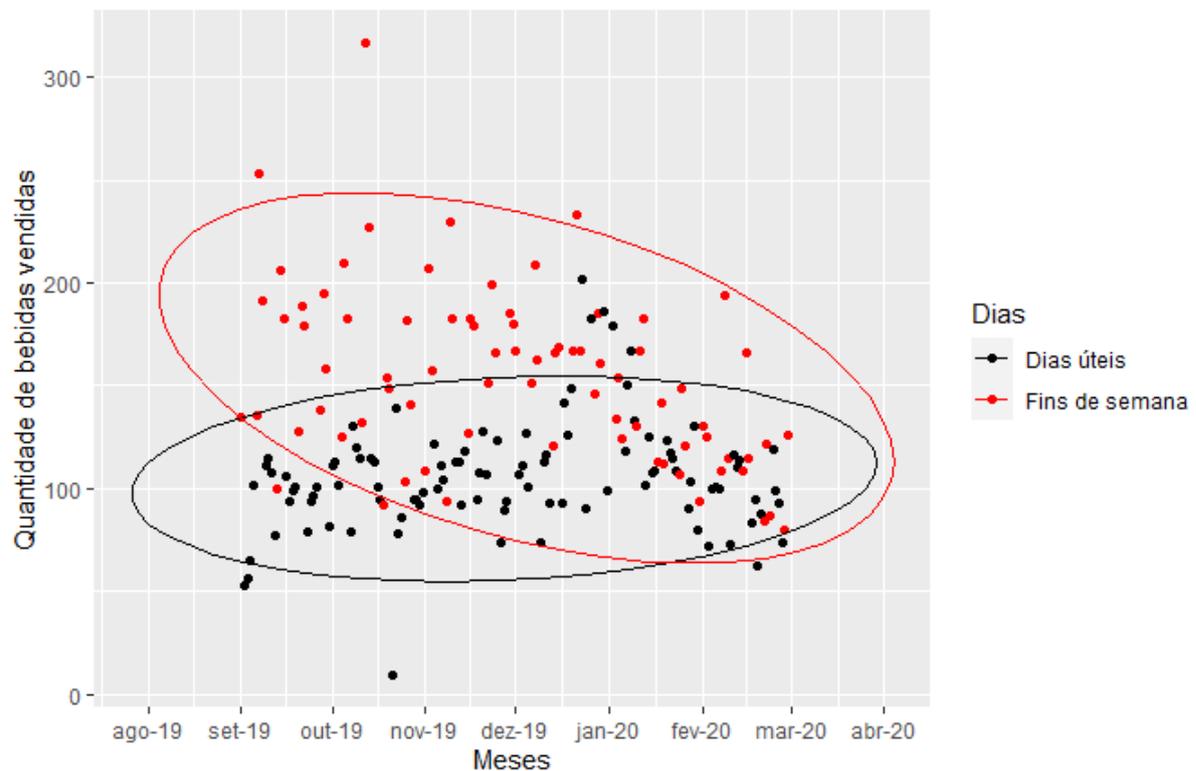
Figura 23 - Estrutura com feriados e fins de semana

| | date | quantity | temp_inst | umid_inst | season | is.weekend | is.holiday |
|---|------------|----------|-----------|-----------|--------|------------|------------|
| 1 | 2019-09-01 | 135 | 16.24286 | 79.64286 | 3 | 1 | 0 |
| 2 | 2019-09-02 | 53 | 16.30000 | 69.30000 | 3 | 0 | 0 |
| 3 | 2019-09-03 | 56 | 19.56500 | 53.85000 | 3 | 0 | 0 |
| 4 | 2019-09-04 | 65 | 16.48095 | 58.23810 | 3 | 0 | 0 |
| 5 | 2019-09-05 | 102 | 18.07059 | 62.70588 | 3 | 0 | 0 |
| 6 | 2019-09-06 | 136 | 20.35714 | 70.71429 | 3 | 1 | 0 |
| 7 | 2019-09-07 | 253 | 25.88636 | 54.09091 | 3 | 1 | 1 |

Fonte: Elaborada pelo autor (2020).

Na Figura 23, podemos confirmar a eficiência da criação desta informação verificando a data de 07 de setembro de 2019, feriado de celebração da independência do Brasil e classificado como um fim de semana, já que em 2019 esta data aconteceu em um sábado. A seguir, vemos uma representação gráfica na Figura 24 que ilustra o aumento de vendas em fins de semana, e a Figura 25, que destaca o aumento de vendas em feriados.

Figura 24 - Vendas em dias úteis e fins de semana

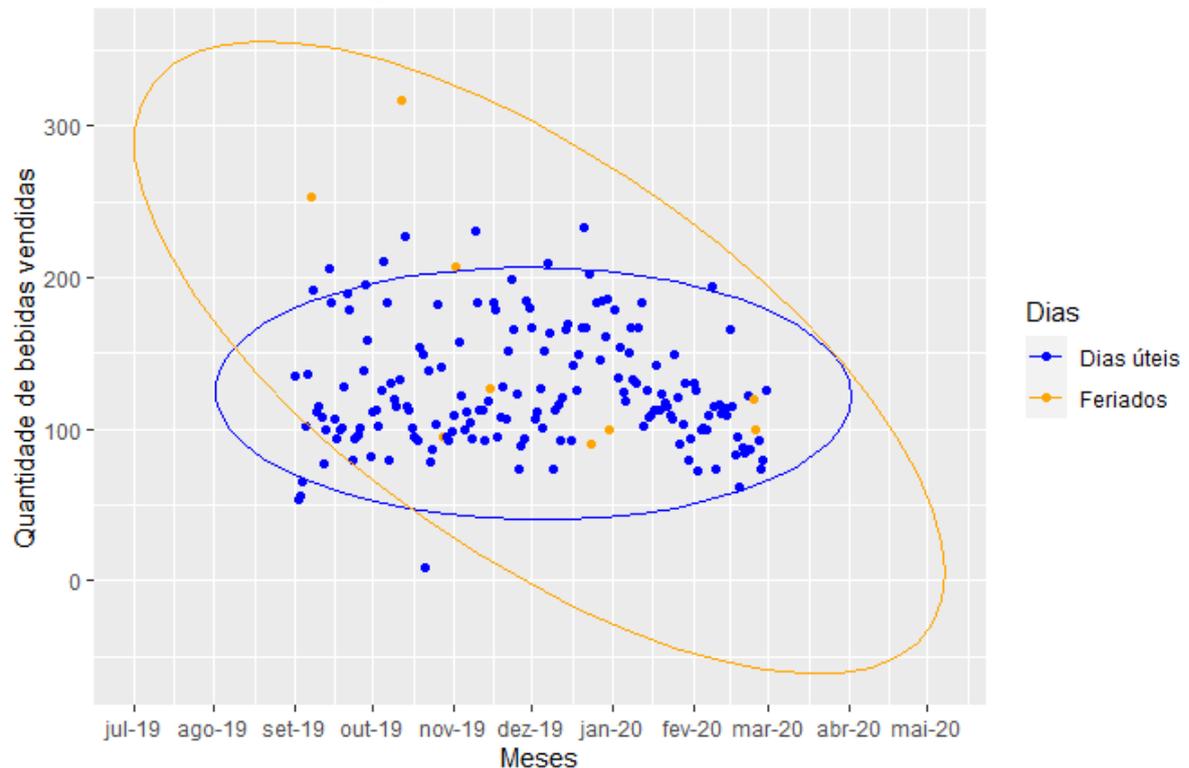


Fonte: Elaborada pelo autor (2020).

Apesar de a Figura 24 demonstrar visualmente que existe uma flutuação positiva que denota o aumento das vendas de *bubble tea* nos fins de semana, o coeficiente de relação de Pearson foi calculado para confirmar matematicamente a existência de uma correlação positiva entre as duas variáveis analisadas, encontrado os valores para $R = 0.57$ e o valor de $p < 0.001$, o que mais uma vez invalida a hipótese nula e confirma uma hipótese alternativa proposta por este estudo. O valor de R deixa explícito que temos uma correlação considerável entre os fins de semana e as vendas nestes dias.

Por se tratar de uma cidade turística amplamente conhecida em território nacional e internacional, a cidade de Foz do Iguaçu possibilita a assunção de uma outra hipótese, de que as vendas podem ser influenciadas positivamente em dias que são categorizados como feriados, por conta da maior probabilidade do aumento do fluxo de pessoas nestes dias, especialmente em feriados prolongados. Para verificar essa hipótese, o gráfico de correlação é ilustrado na Figura 25, em que as vendas em feriados são comparadas com as vendas em dias úteis:

Figura 25 - Vendas em dias úteis e feriados



Fonte: Elaborada pelo autor (2020).

Apesar de existirem pontos em que as vendas atingiram picos maiores em feriados, como no caso dos feriados de 07 de setembro de 12 de outubro, podemos perceber que as vendas em dias úteis em alguns casos ultrapassam as vendas em feriados, observando os pontos dentro do contorno em azul. O coeficiente de relação de Pearson neste caso apresenta os valores para $R = 0.15$ e o valor de $p = 0.032$, o que demonstra que apesar de ter uma significância estatística, a correlação entre as duas variáveis não têm forte tendência, ao contrário do que visualizamos na Figura 24 que trata dos fins de semana.

Por conta da característica do *bubble tea* e seus diferenciais perante produtos concorrentes, a própria franquia objeto de estudo destaca o apelo do produto entre crianças, adolescentes e jovens adultos. Por conta deste apelo, é possível formular a hipótese de que o período de férias escolares pode influenciar na flutuação de vendas dos produtos, pois as pessoas que se encaixam neste perfil, poderiam ter mais oportunidades de visitar o *shopping center* e acabar consumindo o *bubble tea* como uma compra de conveniência. Para investigar se esta hipótese é verdadeira, foi necessário a criação de uma coluna no conjunto de dados com um valor *booleano*,

para verificar se a data das vendas coincide com o período de férias escolares. A Figura 26 demonstra a nova estrutura após acréscimo desta informação.

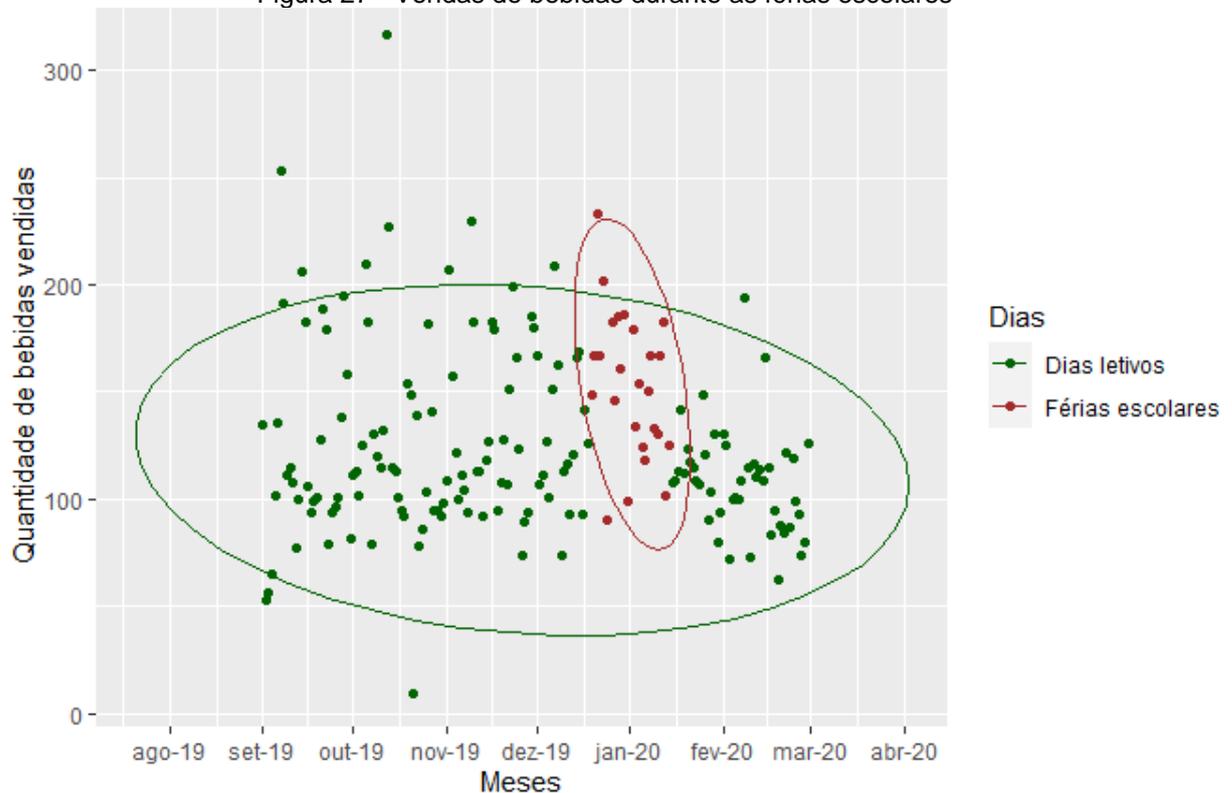
Figura 26 - Estrutura com informações sobre as férias escolares

| | date | quantity | temp_inst | umid_inst | season | is.weekend | is.holiday | is.vacation |
|-----|------------|----------|-----------|-----------|--------|------------|------------|-------------|
| 156 | 2019-12-19 | 149 | 24.34762 | 79.00000 | 4 | 0 | 0 | 1 |
| 157 | 2019-12-20 | 167 | 24.77391 | 83.86957 | 4 | 1 | 0 | 1 |
| 158 | 2019-12-21 | 233 | 22.56190 | 84.47619 | 4 | 1 | 0 | 1 |
| 159 | 2019-12-22 | 167 | 22.06818 | 72.00000 | 4 | 1 | 0 | 1 |
| 160 | 2019-12-23 | 202 | 23.63529 | 56.11765 | 4 | 0 | 0 | 1 |
| 161 | 2019-12-24 | 90 | 22.96364 | 62.86364 | 4 | 0 | 1 | 1 |
| 162 | 2019-12-26 | 183 | 29.32105 | 64.63158 | 4 | 0 | 0 | 1 |
| 163 | 2019-12-27 | 146 | 27.99565 | 71.26087 | 4 | 1 | 0 | 1 |
| 164 | 2019-12-28 | 185 | 28.38750 | 70.91667 | 4 | 1 | 0 | 1 |
| 165 | 2019-12-29 | 161 | 28.05833 | 70.41667 | 4 | 1 | 0 | 1 |

Fonte: Elaborada pelo autor (2020).

Para a definição da faixa de intervalo para as férias escolares foi utilizado o Anexo da Resolução N° 9/2019 da Secretaria de Estado da Educação do Governo do Paraná (Curitiba (PR), 2019). A Figura 27 demonstra a flutuação de vendas durante as férias escolares, comparando-as com as vendas durante período letivo.

Figura 27 - Vendas de bebidas durante as férias escolares



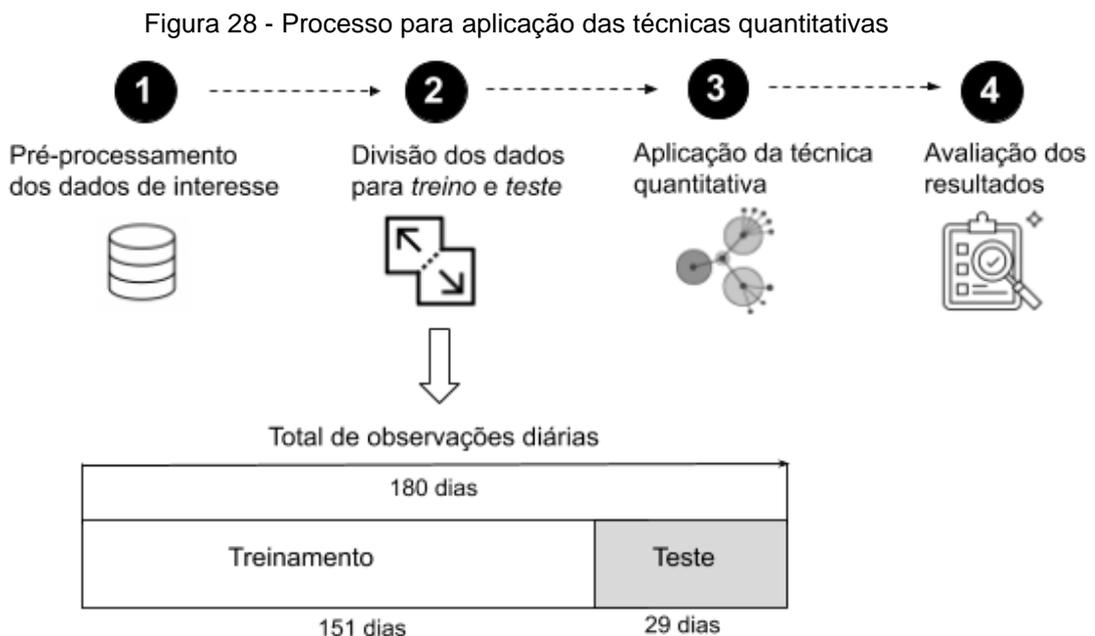
Fonte: Elaborada pelo autor (2020).

Verificando o coeficiente de relação de Pearson temos os valores de $R = 0.25$ e o valor de $p < 0.001$, o que demonstra que existe uma correlação positiva entre a quantidade de bebidas vendidas durante o período de férias escolares, e o valor de p menor do que 0.001 indica a hipótese alternativa que este estudo se baseia, ou seja, de que o período de férias possibilita um aumento nas vendas.

Utilizando as informações descobertas na análise exploratória dos dados e do incremento no conjunto de dados para contemplar estas descobertas, na Seção 4.1 testes com as técnicas quantitativas previamente detalhadas no Capítulo 3 serão apresentados e discutidos.

4.2. Aplicação das Técnicas Quantitativas selecionadas

Após análise exploratória acerca do conjunto de dados criado para este estudo, as técnicas quantitativas discutidas no Capítulo 3 foram aplicadas com o objetivo de identificar o modelo mais adequado para a proposta de modelo preditivo de vendas para a franquia Bubble Mix. A Figura 28 ilustra o processo para aplicação das técnicas quantitativas apresentadas nesta pesquisa.



Fonte: Elaborada pelo autor (2020).

Os dados de interesse foram pré-processados para abarcar as variáveis descobertas durante a análise experimental e o conjunto de dados foi dividido em duas partes: treino e teste. O primeiro consiste em 151 observações diárias entre 01 de setembro de 2019 a 31 de janeiro de 2020 e o segundo em 29 observações diárias entre 01 e 29 de fevereiro de 2020. Estas observações contemplam os produtos vendidos durante o período. Após a separação dos dados, a técnica quantitativa a ser testada foi aplicada e uma avaliação dos resultados foi feita utilizando os dados de teste. O objetivo desta etapa foi verificar os erros de previsão para cada técnica utilizando as seguintes métricas: EAM (Erro Absoluto Médio), Erro Percentual Absoluto Médio (EPAM) e Raiz do Erro Quadrático Médio (REQM). Nas próximas seções, serão verificados os erros para as técnicas de Médias Móveis, Suavização Exponencial e Box-Jenkins (ARIMA).

4.2.1. Aplicação da técnica de Médias Móveis

As médias móveis realizam uma tarefa de suavização dos dados relativos a vendas para construção de um indicador de tendência da sequência observada. O objetivo não se torna prever a direção das vendas, mas definir a direção atual com um atraso baseado nos dados passados, filtrando seus ruídos. Utilizando o processo detalhado na Figura 28, a técnica de Médias Móveis foi aplicada criando uma previsão de vendas para o mês de fevereiro de 2020.

A Tabela 4 exhibe os valores, comparando as vendas reais com a previsão, calculando a margem de erro diária em percentual e a Figura 29 ilustra o resultado obtido.

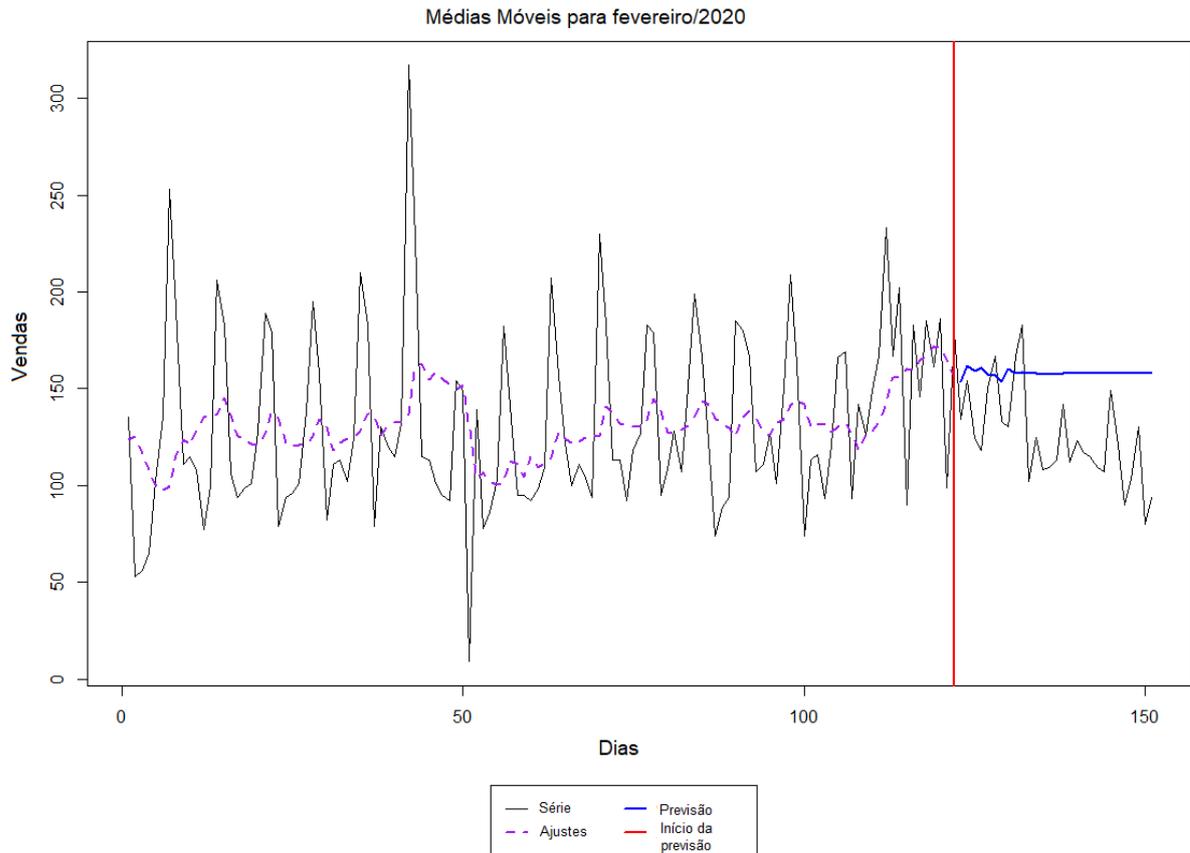
Tabela 4 - Cálculo de Médias Móveis para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 154 | 18% |
| 02/02/2020 | 125 | 162 | 30% |
| 03/02/2020 | 72 | 159 | 121% |
| 04/02/2020 | 100 | 161 | 61% |
| 05/02/2020 | 101 | 157 | 55% |
| 06/02/2020 | 100 | 157 | 57% |
| 07/02/2020 | 109 | 153 | 40% |
| 08/02/2020 | 194 | 160 | -18% |
| 09/02/2020 | 115 | 158 | 37% |
| 10/02/2020 | 73 | 158 | 116% |
| 11/02/2020 | 116 | 158 | 36% |
| 12/02/2020 | 110 | 158 | 44% |
| 13/02/2020 | 114 | 158 | 39% |
| 14/02/2020 | 109 | 158 | 45% |
| 15/02/2020 | 166 | 158 | -5% |
| 16/02/2020 | 115 | 158 | 37% |
| 17/02/2020 | 83 | 158 | 90% |
| 18/02/2020 | 95 | 158 | 66% |
| 19/02/2020 | 62 | 158 | 155% |
| 20/02/2020 | 88 | 158 | 80% |
| 21/02/2020 | 84 | 158 | 88% |
| 22/02/2020 | 122 | 158 | 30% |
| 23/02/2020 | 87 | 158 | 82% |
| 24/02/2020 | 119 | 158 | 33% |
| 25/02/2020 | 99 | 158 | 60% |
| 26/02/2020 | 93 | 158 | 70% |
| 27/02/2020 | 74 | 158 | 114% |
| 28/02/2020 | 80 | 158 | 98% |
| 29/02/2020 | 126 | 158 | 25% |

Fonte: Elaborada pelo autor (2020).

Na Figura 29 temos a série contemplando todas as observações do conjunto de dados (180 dias), os valores ajustados pelo modelo ilustradas pela linha pontilhada, o início da previsão (a partir de observação 151, conforme os dados separados para treinamento) e a partir da observação 151 em diante temos a linha azul ilustrando as previsões pelas médias móveis. A linha da série demonstra o comportamento real da série (dados para teste), sendo possível comparar visualmente o resultado do modelo com os resultados reais. A Tabela 5 exibe as métricas de erro encontrado às médias móveis considerando o conjunto de teste para efeitos comparativos.

Figura 29 - Médias Móveis para fevereiro/2020



Fonte: Elaborada pelo autor (2020).

Tabela 5 - Métricas para erros de previsão de Médias Móveis

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 51.75 |
| Erro Percentual Absoluto Médio (MAPE) | 56.33 |
| Raiz Quadrada do Erro Médio (RMSE) | 55.20 |

Fonte: Elaborada pelo autor (2020).

Pela Tabela 5 podemos identificar que houve um EAM de aproximadamente 56 unidades vendidas neste modelo de previsão, sendo o EPAM de 56,33% e o REQM de 55.20, sendo este último valor considerado alto pois quando mais próximo de zero, significa que melhor é a previsão. Pela simplicidade de sua fórmula, as médias móveis apresentam uma forma de previsão com uma margem de erro considerada alta, pois não leva em consideração tendência, sazonalidade ou pontos de virada em suas observações.

4.2.2. Aplicação da técnica de Suavização Exponencial

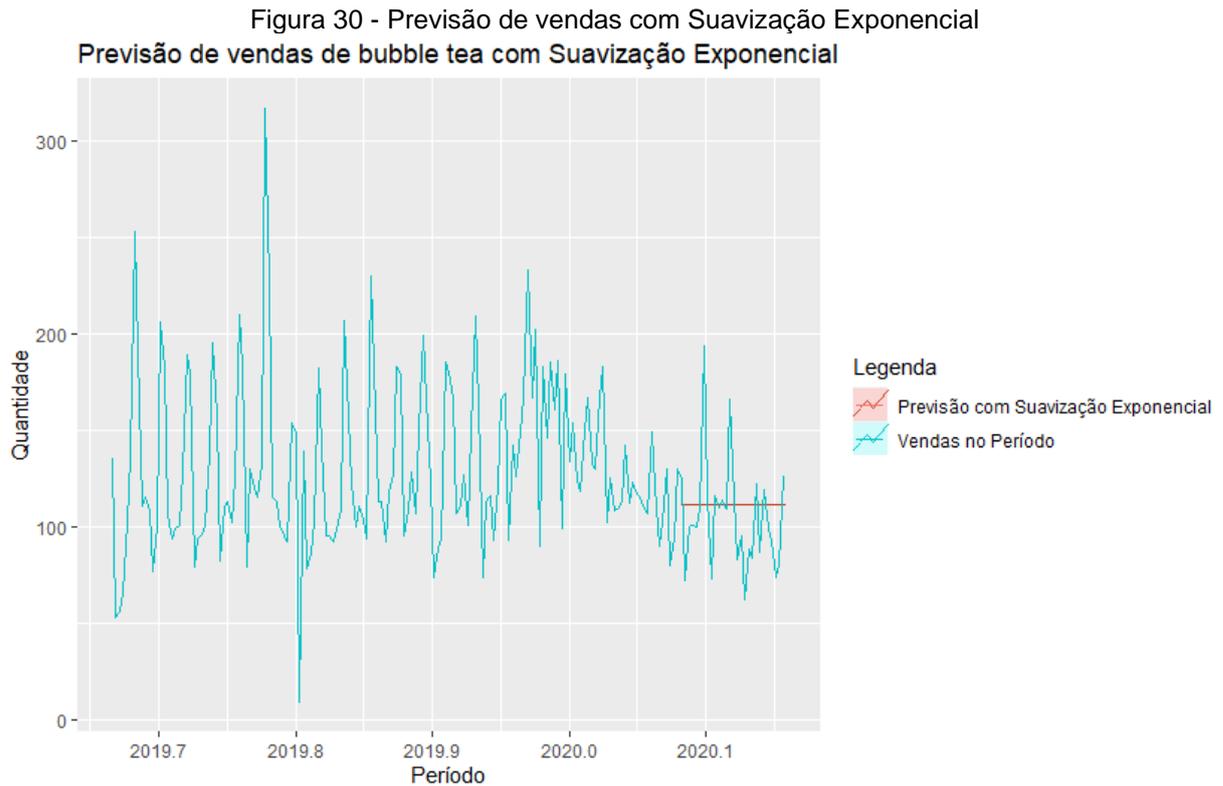
Enquanto as Médias Móveis atribuem pesos iguais para os valores distribuídos no tempo, a Suavização Exponencial permite que as médias sejam ponderadas em que pesos maiores podem ser atribuídos a observações mais recentes, e menores pesos em observações mais antigas. Esse método é considerado bastante intuitivo, computacionalmente eficiente e geralmente aplicável a uma grande variedade de séries temporais. Por meio dos processos descritos na Figura 28, uma previsão foi gerada utilizando a técnica de suavização exponencial. A Tabela 6 exibe os valores encontrados por essa projeção.

Tabela 6 - Cálculo de Médias Móveis para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 111 | -15% |
| 02/02/2020 | 125 | 111 | -11% |
| 03/02/2020 | 72 | 111 | 54% |
| 04/02/2020 | 100 | 111 | 11% |
| 05/02/2020 | 101 | 111 | 10% |
| 06/02/2020 | 100 | 111 | 11% |
| 07/02/2020 | 109 | 111 | 2% |
| 08/02/2020 | 194 | 111 | -43% |
| 09/02/2020 | 115 | 111 | -3% |
| 10/02/2020 | 73 | 111 | 52% |
| 11/02/2020 | 116 | 111 | -4% |
| 12/02/2020 | 110 | 111 | 1% |
| 13/02/2020 | 114 | 111 | -3% |
| 14/02/2020 | 109 | 111 | 2% |
| 15/02/2020 | 166 | 111 | -33% |
| 16/02/2020 | 115 | 111 | -3% |
| 17/02/2020 | 83 | 111 | 34% |
| 18/02/2020 | 95 | 111 | 17% |
| 19/02/2020 | 62 | 111 | 79% |
| 20/02/2020 | 88 | 111 | 26% |
| 21/02/2020 | 84 | 111 | 32% |
| 22/02/2020 | 122 | 111 | -9% |
| 23/02/2020 | 87 | 111 | 28% |
| 24/02/2020 | 119 | 111 | -7% |
| 25/02/2020 | 99 | 111 | 12% |
| 26/02/2020 | 93 | 111 | 19% |
| 27/02/2020 | 74 | 111 | 50% |
| 28/02/2020 | 80 | 111 | 39% |
| 29/02/2020 | 126 | 111 | -12% |

Fonte: Elaborada pelo autor (2020).

A Figura 30 demonstra o comportamento da série, com a linha em azul ilustrando a flutuação de vendas no período de setembro de 2019 a fevereiro de 2020. A linha em vermelho ilustra a previsão de vendas com a técnica de Suavização Exponencial. Fica evidente a partir deste ponto, que a limitação dos dados disponíveis influencia diretamente na qualidade da previsão a ser realizada.



Fonte: Elaborada pelo autor (2020).

Apesar dos dados históricos utilizados para treinamento do modelo, ao realizar uma previsão de um horizonte de 29 dias (mês de fevereiro de 2020), a linha se mantém reta por todo o período, demonstrando que mesmo com a utilização do valor de $\alpha = 0.2$, a série se mantém constante pois a previsão para esses dias é feita com observações passadas, mas sem a atualização das vendas mais recentes, que poderiam ser utilizadas para melhorar o modelo. Essa limitação se deve ao fato de a massa de dados disponibilizada para esta pesquisa não contemplar um período maior, por conta da diferença de formato dos dados do sistema utilizado anteriormente e da janela temporal disponibilizada pela franqueadora para a pesquisa. A Tabela 7 exibe as métricas de erros encontradas para esse modelo.

Tabela 7 - Métricas para erros de previsão da Suavização Exponencial

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 20.68 |
| Erro Percentual Absoluto Médio (MAPE) | 21.43 |
| Raiz Quadrada do Erro Médio (RMSE) | 27.75 |

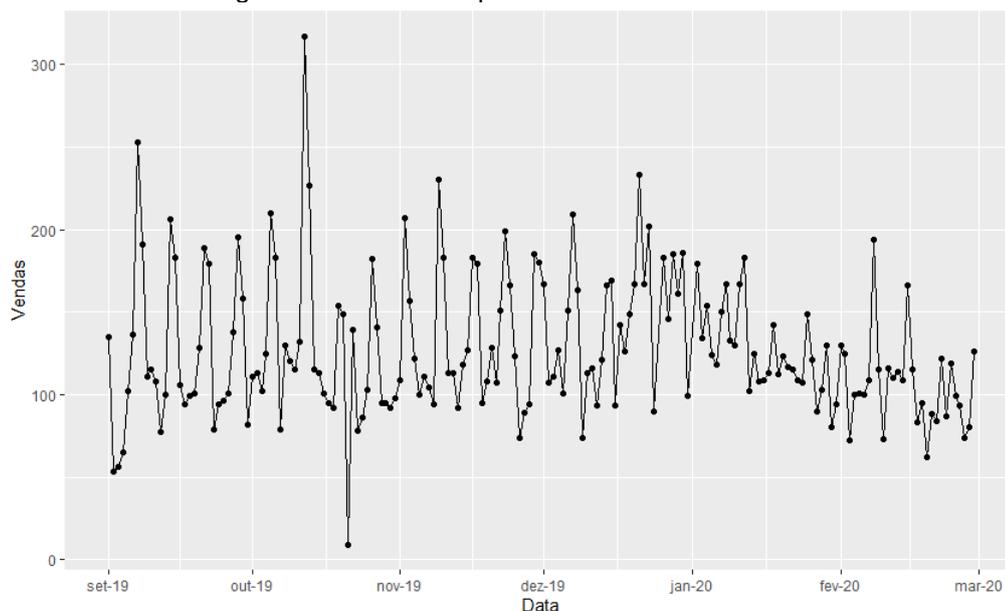
Fonte: Elaborada pelo autor (2020).

A Tabela 7 demonstra um EAM de aproximadamente 20 bebidas considerando todo o conjunto de dados, com um EPAM de 21,43% e um REQM de 27.75. Podemos observar que embora as métricas sejam melhores do que as médias móveis, a suavização exponencial ainda possui limitações relacionadas a sua capacidade de previsão, por conta da simplicidade do modelo.

4.2.3. Aplicação do modelo ARIMA

Para uma identificação apropriada do modelo, o primeiro passo é analisar o gráfico de séries temporais das observações pois este pode indicar uma tendência ou uma alteração em sua flutuação, o que pode contribuir para identificar se uma série é estacionária ou não estacionária. A Figura 31 demonstra a série temporal observada durante o período de setembro de 2019 a março de 2020.

Figura 31 - Séries temporais vendas de *bubble tea*



Fonte: Elaborada pelo autor (2020).

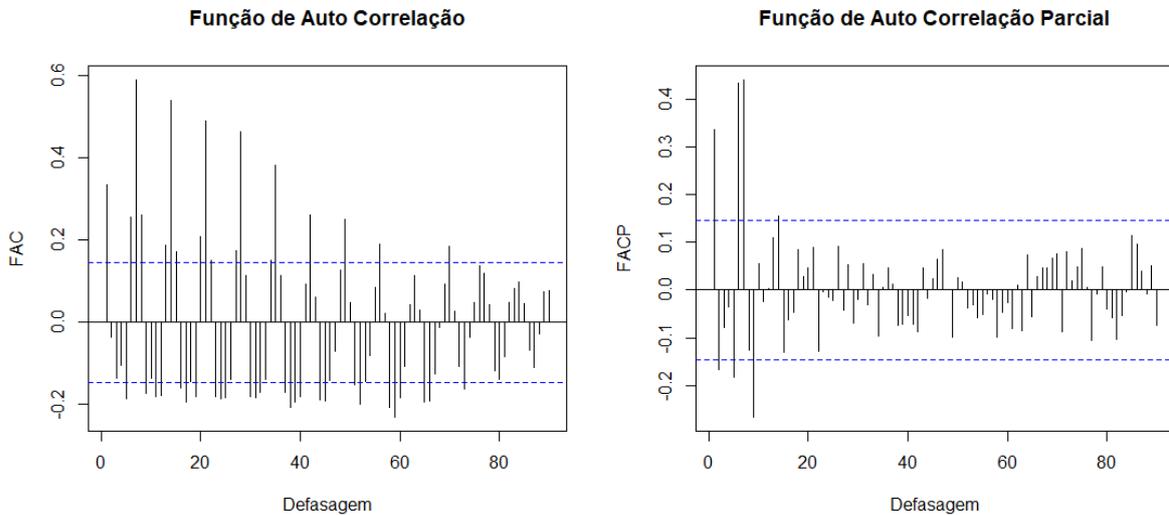
Observando a Figura 31 é possível afirmar que a série apresenta uma leve tendência decrescente e com pouca inclinação, o que fortalece a hipótese se tratar de uma série estacionária. Para a confirmação desta hipótese, aplicamos o teste de Dickey-Fuller com o resultado demonstrado pela Figura 32, que confirma a hipótese de que se trata de uma série estacionária. Por conta desta confirmação, não é necessário realizar o processo de diferenciação na série e ela está apta a comportar uma simulação utilizando ARIMA. O próximo passo seria a identificação do modelo a ser utilizado e, portanto, torna-se necessário visualizar a Função de Auto Correlação (FAC) e a Função de Auto Correlação Parcial (FACP) para as séries observadas, limitando estas as primeiras 90 observações, conforme visualização da Figura 33.

Figura 32 - Resultado do teste de Dickey-Fuller
Augmented Dickey-Fuller Test

```
data: ts_sales
Dickey-Fuller = -4.4602, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Fonte: Função `adf.test()` do pacote *tseries* (Trapletti & Hornik, 2019).

Figura 33 – FAC e FACP



Fonte: Elaborada pelo autor (2020).

Analisando a FAC ilustrada na Figura 33, vemos que a série inicia com uma correlação negativa nas 4 primeiras observações e logo após vemos picos acima de zero nas três observações seguintes, voltando rapidamente para zero o que ilustra que não há correlação forte entre as observações passadas e as futuras. Séries temporais que demonstram não ter correlação são chamadas de ruído branco. Para estas séries é esperado que cada correlação seja próxima de zero mesmo que haja uma pequena variação aleatória (HYNDMAN & ATHANASOPOULOS, 2018).

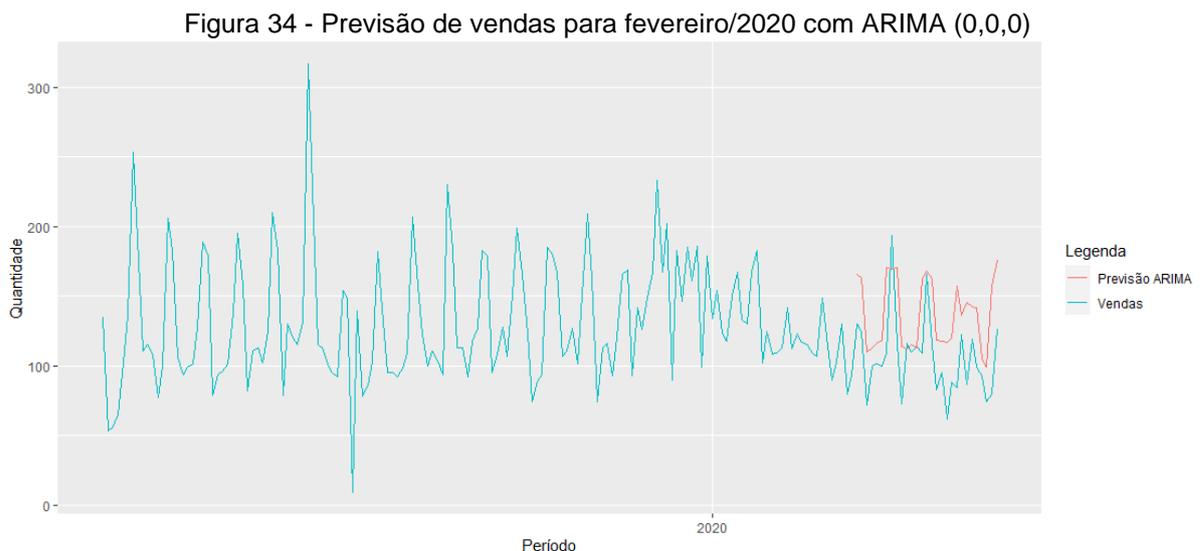
Para identificação dos elementos ARIMA(p, d, q), o modelo será escolhido utilizando o critério de AIC (*Akaike Information Criteria*), pois para Aho, Derryberry & Peterson (2014, p. 633): “Critério como o AIC maximiza a precisão preditiva”. Utilizando a função *auto.arima()*, do pacote *forecast* (HYNDMAN *et al.*,2020) da linguagem de programação R, temos como retorno alguns modelos avaliados e a definição do melhor modelo a ser utilizado, conforme detalhes exibidos na Tabela 8.

Tabela 8 – Critério para seleção do modelo

| Modelo | AIC |
|----------------|----------|
| ARIMA(2, 0, 2) | 1496.219 |
| ARIMA(0, 0, 0) | 1492.354 |
| ARIMA(1, 0, 0) | 1495.432 |
| ARIMA(0, 0, 1) | 1494.485 |
| ARIMA(1, 0, 1) | 1494.222 |

Fonte: Elaborada pelo autor (2020).

Segundo o critério AIC, observando o valor de 1492.35 na Tabela 8, encontramos o modelo ARIMA(0, 0, 0) como sendo o melhor modelo. Após aplicação do modelo selecionado, obtemos a previsão para o mês de fevereiro conforme a Tabela 9. Na Figura 34, a linha azul representa a flutuação de vendas durante o período observado, tendo a linha laranja como sendo a previsão realizada pelo modelo ARIMA (0, 0, 0).



Fonte: Elaborada pelo autor (2020).

Para verificar a acurácia do modelo, a Tabela 10 demonstra as métricas de erro observadas.

Tabela 9 – Previsão ARIMA (0, 0, 0) para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 166 | 28% |
| 02/02/2020 | 125 | 163 | 30% |
| 03/02/2020 | 72 | 110 | 53% |
| 04/02/2020 | 100 | 113 | 13% |
| 05/02/2020 | 101 | 116 | 15% |
| 06/02/2020 | 100 | 118 | 18% |
| 07/02/2020 | 109 | 170 | 56% |
| 08/02/2020 | 194 | 170 | -12% |
| 09/02/2020 | 115 | 170 | 48% |
| 10/02/2020 | 73 | 114 | 56% |
| 11/02/2020 | 116 | 112 | -3% |
| 12/02/2020 | 110 | 115 | 5% |
| 13/02/2020 | 114 | 113 | -1% |
| 14/02/2020 | 109 | 163 | 50% |
| 15/02/2020 | 166 | 168 | 1% |
| 16/02/2020 | 115 | 162 | 41% |
| 17/02/2020 | 83 | 118 | 42% |
| 18/02/2020 | 95 | 118 | 24% |
| 19/02/2020 | 62 | 116 | 87% |
| 20/02/2020 | 88 | 120 | 36% |
| 21/02/2020 | 84 | 157 | 87% |
| 22/02/2020 | 122 | 136 | 11% |
| 23/02/2020 | 87 | 146 | 68% |
| 24/02/2020 | 119 | 142 | 19% |
| 25/02/2020 | 99 | 141 | 42% |
| 26/02/2020 | 93 | 106 | 14% |
| 27/02/2020 | 74 | 99 | 34% |
| 28/02/2020 | 80 | 157 | 96% |
| 29/02/2020 | 126 | 176 | 40% |

Fonte: Elaborada pelo autor (2020).

Tabela 10 - Métricas para erros de previsão ARIMA (0, 0, 0)

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 33.65 |
| Erro Percentual Absoluto Médio (MAPE) | 35.55 |
| Raiz Quadrada do Erro Médio (RMSE) | 39.64 |

Fonte: Elaborada pelo autor (2020).

Neste modelo torna-se ainda mais relevante a limitação da amostra de dados disponíveis, pois a ausência de um ciclo completo de vendas (período de 365 dias), torna-se impossível verificar com eficiência os efeitos sobre a sazonalidade na flutuação das vendas. O modelo poderia ter um resultado melhor com uma janela de

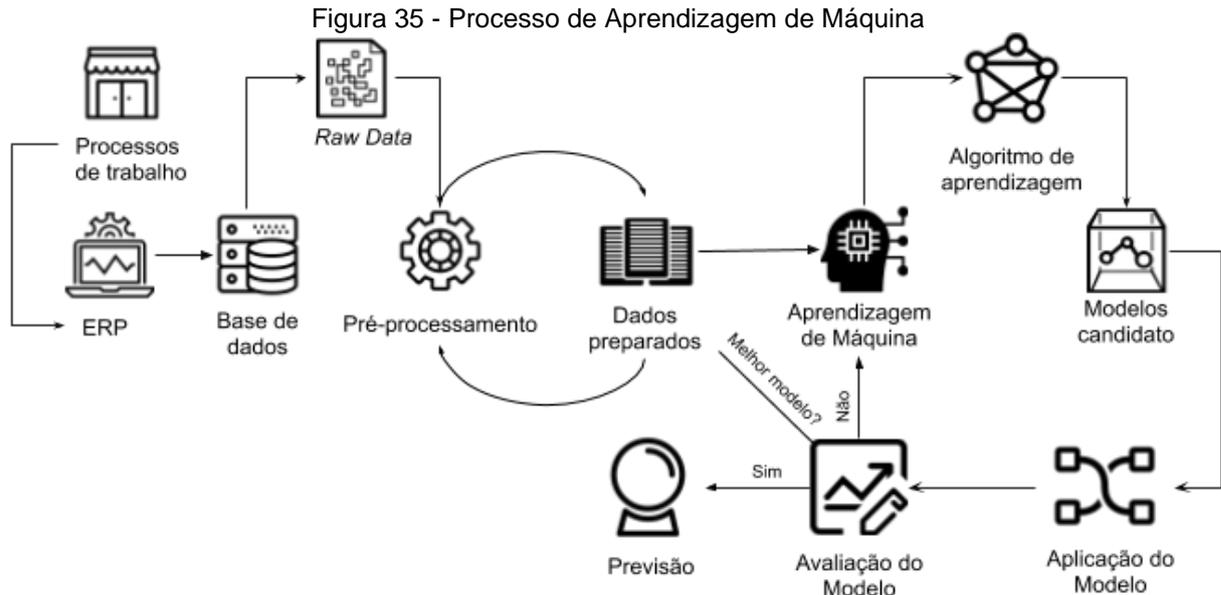
tempo maior, contendo mais observações para alimentar os componentes auto regressivos. Observando a Tabela 10, temos o valor para EAM de aproximadamente 33 bebidas, com um EPAM de 35,55% e um REQM de 39,64 demonstram que o modelo é mais preciso que Médias Móveis, mas com métricas inferiores se comparado com a Suavização Exponencial.

Observando a Figura 34, vemos que a faixa de previsão apresenta uma flutuação que reflete a inconstância nas vendas. Por conta desta limitação, utilizaremos métodos baseados em inteligência computacional para comparar a sua acurácia com os métodos quantitativos.

4.2.4. Aplicação das Técnicas de Aprendizagem de Máquina

Para que seja possível que o algoritmo aprenda sobre os dados, são passados dados de entrada separados como dados de “treinamento” e então, seu desempenho é avaliado a partir de dados “desconhecidos”, separados como dados de “teste”. A aplicação das Técnicas de Aprendizagem de Máquina utiliza a mesma separação do conjunto de dados para as etapas de treinamento e teste ilustradas na Figura 28. Os erros gerados pelo algoritmo durante a sua fase de treinamento, nos dão a ideia de como esse algoritmo está se comportando com relação aos dados analisados. Estes erros também podem ajudar a esclarecer se o algoritmo está usando sua capacidade plena ou se necessita de ajustes em seus parâmetros (GHATAK, 2017).

A Figura 35 ilustra o processo de aplicação de Técnicas de Aprendizagem de Máquina, na qual podemos observar que o contexto denota os fatores ambientais da empresa, ou seja, sua área de atuação e as rotinas diárias de trabalho. As saídas dos processos de trabalho, alimentam seu sistema ERP, que por sua vez, armazena os dados em um banco de dados. Porém, caso deste estudo, esses dados são classificados como *raw data*, ou seja, dados “crus”, apresentados em formatos diferentes do necessário para aplicação das diferentes técnicas.



Esses dados foram limpos e tiveram sua integridade verificada por meio de atividades de pré-processamento até que estivessem prontos para o início do processo de aprendizagem. Então, para cada método selecionado: Redes Neurais Artificial *feedforward*, Modelos Lineares Generalizados e Florestas Aleatórias, seu respectivo algoritmo àquele conjunto de dados preparados, indicou um ou mais modelos candidatos, que foram aplicados aos dados possibilitando a avaliação do modelo. Caso o modelo avaliado fosse compatível com o problema a ser resolvido, uma previsão era então gerada, caso contrário, o ciclo iterativo se reiniciaria até que um modelo satisfatório fosse encontrado.

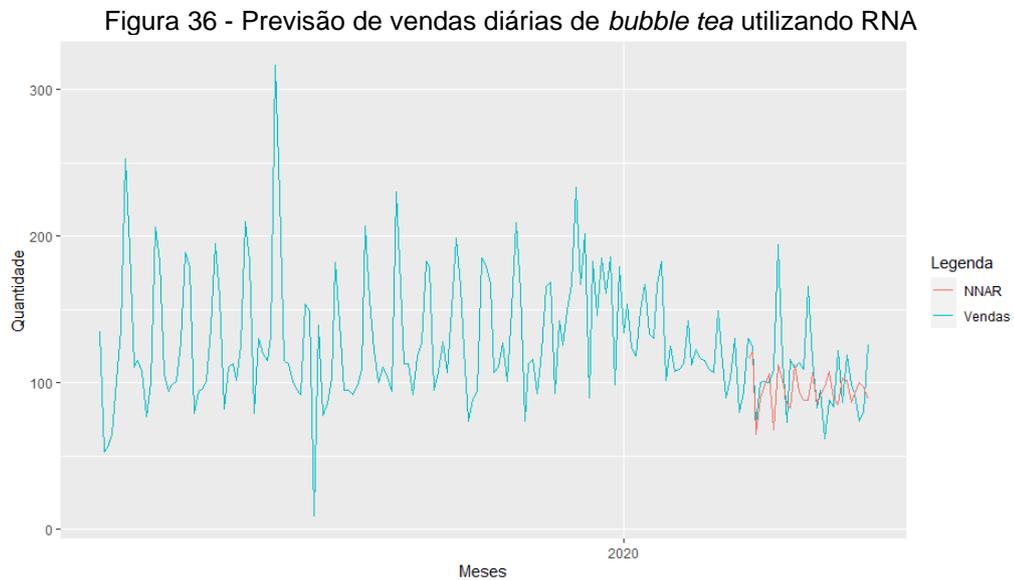
4.2.5. Aplicação de Redes Neurais Artificiais

Zhang, Patuwo & Hu (1998) afirmam que as redes MLP são utilizadas em uma grande variedade de problemas, especialmente em tarefas de previsão por sua capacidade inerente de mapeamento arbitrário de entradas e saídas. Como temos os dados das séries temporais observadas, podemos utilizar os valores defasados das séries temporais como entradas para uma RNA, e com isso, usaremos um modelo de auto regressão de RNA conhecida como NNAR (*Neural Network Auto Regressive*),

por meio da função *nnetar()* do pacote *forecast* (HYNDMAN *et al.*, 2019) presente na linguagem de programação R.

Utilizando os dados de treinamento como entrada para a função, o melhor modelo encontrado foi o NNAR(9, 5) que representa um modelo com as nove últimas observações usadas como entrada para previsão do valor alvo e cinco neurônios em sua camada oculta. A Tabela 11 exibe os valores encontrados para a previsão.

A Figura 36 exibe as vendas observadas de setembro de 2019 a fevereiro de 2020, representadas pela linha na cor azul, enquanto a linha laranja representa a previsão realizada utilizando o modelo NNAR (9, 5).



Fonte: Elaborada pelo autor (2020).

Tabela 11 – Previsão NNAR(9, 5) para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 116 | -11% |
| 02/02/2020 | 125 | 121 | -3% |
| 03/02/2020 | 72 | 65 | -10% |
| 04/02/2020 | 100 | 90 | -10% |
| 05/02/2020 | 101 | 99 | -2% |
| 06/02/2020 | 100 | 106 | 6% |
| 07/02/2020 | 109 | 68 | -38% |
| 08/02/2020 | 194 | 112 | -42% |
| 09/02/2020 | 115 | 102 | -11% |
| 10/02/2020 | 73 | 86 | 18% |
| 11/02/2020 | 116 | 83 | -28% |
| 12/02/2020 | 110 | 112 | 2% |
| 13/02/2020 | 114 | 94 | -18% |
| 14/02/2020 | 109 | 88 | -19% |
| 15/02/2020 | 166 | 88 | -47% |
| 16/02/2020 | 115 | 107 | -7% |
| 17/02/2020 | 83 | 86 | 4% |
| 18/02/2020 | 95 | 91 | -4% |
| 19/02/2020 | 62 | 97 | 56% |
| 20/02/2020 | 88 | 108 | 23% |
| 21/02/2020 | 84 | 89 | 6% |
| 22/02/2020 | 122 | 85 | -30% |
| 23/02/2020 | 87 | 103 | 18% |
| 24/02/2020 | 119 | 102 | -14% |
| 25/02/2020 | 99 | 87 | -12% |
| 26/02/2020 | 93 | 93 | 0% |
| 27/02/2020 | 74 | 100 | 35% |
| 28/02/2020 | 80 | 97 | 21% |
| 29/02/2020 | 126 | 89 | -29% |

Fonte: Elaborada pelo autor (2020).

As métricas para os erros de previsão são apresentadas na Tabela 12.

Tabela 12 - Métricas para erros de previsão NNAR (9, 5)

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 20.10 |
| Erro Percentual Absoluto Médio (MAPE) | 18.12 |
| Raiz Quadrada do Erro Médio (RMSE) | 28.34 |

Fonte: Elaborada pelo autor (2020).

Observando as métricas para erros de previsão, até o momento a utilização de RNA para previsões apresentou o menor MAPE entre todas as previsões até o momento. Porém, por conta das limitações dos dados disponíveis para a etapa de aprendizagem, ainda é possível notar o MAE com um valor relativamente alto, assim

como o RMSE. Mas se comparadas com as técnicas quantitativas, a Rede Neural Artificial demonstrou maior potencial para realizar previsões.

4.2.6. Aplicação de Modelos Lineares Generalizados

A partir de utilização de técnicas mais avançadas de Aprendizagem de Máquina, torna-se possível utilizar recursos computacionais para processar e avaliar modelos mais complexos que possam ser utilizados para predição de estoques. Tanto para os Modelos Lineares Generalizados (MLG), quanto para o modelo da próxima seção, as Florestas Aleatórias, o conjunto de dados foi expandido para levar em consideração não só as vendas diárias como um todo, mas cada venda específica de acordo com a categoria da bebida. Essa facilidade foi permitida pela utilização do pacote *caret* criado por Max Kuhn (2020).

O pacote *caret* permite a utilização da função `train()` que utiliza um algoritmo de aprendizagem com a possibilidade da escolha do método e a configuração de parâmetros para controle do treinamento. O pacote também fornece as métricas para avaliar a acurácia das predições: MAE, RMSE e R^2 .

Para fins comparativos, a Tabela 13 exibe a previsão realizada pelo modelo MLG.

Tabela 13 – Previsão MLG para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 165 | -27% |
| 02/02/2020 | 125 | 126 | -1% |
| 03/02/2020 | 72 | 81 | -12% |
| 04/02/2020 | 100 | 99 | 1% |
| 05/02/2020 | 101 | 101 | 0% |
| 06/02/2020 | 100 | 140 | -40% |
| 07/02/2020 | 109 | 124 | -14% |
| 08/02/2020 | 194 | 192 | 1% |
| 09/02/2020 | 115 | 139 | -21% |
| 10/02/2020 | 73 | 92 | -26% |
| 11/02/2020 | 116 | 85 | 27% |
| 12/02/2020 | 110 | 108 | 2% |
| 13/02/2020 | 114 | 99 | 13% |
| 14/02/2020 | 109 | 109 | 0% |
| 15/02/2020 | 166 | 179 | -8% |
| 16/02/2020 | 115 | 121 | -5% |
| 17/02/2020 | 83 | 97 | -17% |
| 18/02/2020 | 95 | 88 | 7% |
| 19/02/2020 | 62 | 81 | -31% |
| 20/02/2020 | 88 | 104 | -18% |
| 21/02/2020 | 84 | 105 | -25% |
| 22/02/2020 | 122 | 151 | -24% |
| 23/02/2020 | 87 | 144 | -66% |
| 24/02/2020 | 119 | 114 | 4% |
| 25/02/2020 | 99 | 103 | -4% |
| 26/02/2020 | 93 | 96 | -3% |
| 27/02/2020 | 74 | 79 | -7% |
| 28/02/2020 | 80 | 81 | -1% |
| 29/02/2020 | 126 | 153 | -21% |

Fonte: Elaborada pelo autor (2020).

Na Tabela 14 podemos observar as métricas de erros de previsão para os produtos agrupados, ou seja, considerando as vendas diárias sem distinção por categoria ou produto.

Tabela 14 - Métricas para erros de previsão utilizando MLG

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 14.51 |
| Erro Percentual Absoluto Médio (MAPE) | 14.68 |
| Raiz Quadrada do Erro Médio (RMSE) | 20.05 |

Fonte: Elaborada pelo autor (2020).

O modelo implementado para este caso, oferta a possibilidade de retornar a previsão por dia e por produto, levando em consideração a sua categoria e os

preditores disponíveis para o modelo. A Figura 37 demonstra um recorte do conjunto de dados gerados pelo MLG e a Figura 38 ilustra a previsão de vendas por dia, onde temos a linha em azul como sendo os dados observados e a linha em laranja os valores previstos para o mês de fevereiro de 2020.

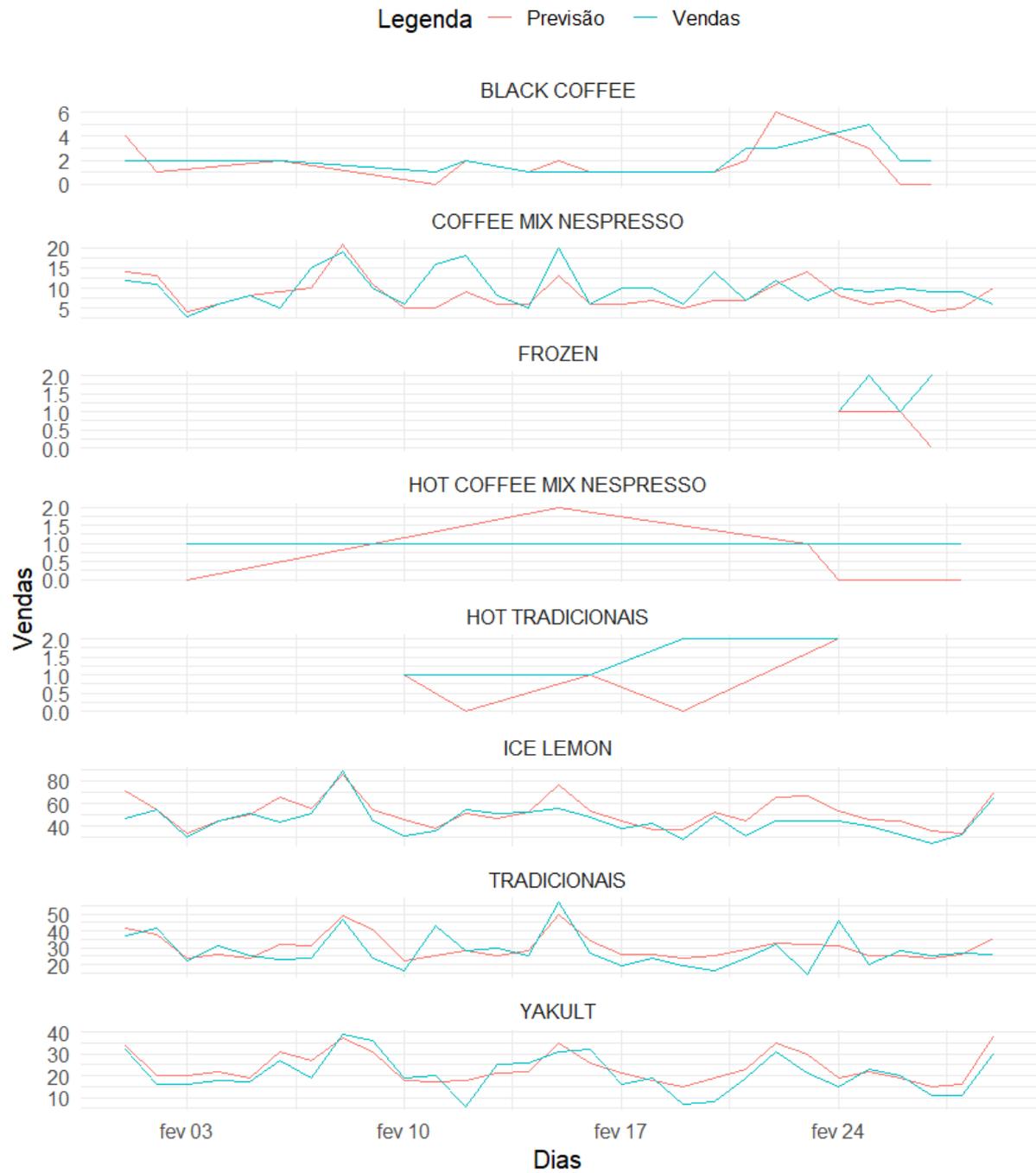
Figura 37 - Recorte do conjunto de dados gerado pelo MLG

| product | quantity | date | day | weekday | is.holiday | is.vacation | is.weekend | category | temp_inst | umid_inst | season | predicted | error |
|---------------------------------------|----------|------------|-----|---------|------------|-------------|------------|--------------------|-----------|-----------|--------|-----------|-------|
| BLACK COFFEE - ESPRESSO FORTE | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | BLACK COFFEE | 26.30455 | 72.13636 | 4 | 2 | 1 |
| BLACK COFFEE - LUNGO FORTE | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | BLACK COFFEE | 26.30455 | 72.13636 | 4 | 2 | 1 |
| COFFEE MIX NESPRESSO - AMARETTO | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | COFFEE MIX NESP... | 26.30455 | 72.13636 | 4 | 2 | 1 |
| TRADICIONAIS - ASSAM CREAMY | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | TRADICIONAIS | 26.30455 | 72.13636 | 4 | 2 | 1 |
| ICE LEMON - BERRY GO | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | ICE LEMON | 26.30455 | 72.13636 | 4 | 2 | 0 |
| ICE LEMON - FRESH'N GINGER | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | ICE LEMON | 26.30455 | 72.13636 | 4 | 2 | 0 |
| TRADICIONAIS - INDIAN CHAI | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | TRADICIONAIS | 26.30455 | 72.13636 | 4 | 2 | 0 |
| YAKULT - NATURAL | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | YAKULT | 26.30455 | 72.13636 | 4 | 2 | 0 |
| ICE LEMON - CLASSIC PEACH | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | ICE LEMON | 26.30455 | 72.13636 | 4 | 3 | 2 |
| TRADICIONAIS - COCONUT CREAMY | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | TRADICIONAIS | 26.30455 | 72.13636 | 4 | 3 | 2 |
| YAKULT - CRANBERRY | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | YAKULT | 26.30455 | 72.13636 | 4 | 3 | 2 |
| YAKULT - MELÃO | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | YAKULT | 26.30455 | 72.13636 | 4 | 3 | 2 |
| YAKULT - LIMÃO SICILIANO | 1 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | YAKULT | 26.30455 | 72.13636 | 4 | 3 | 2 |
| COFFEE MIX NESPRESSO - DULCE DE LECHE | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | COFFEE MIX NESP... | 26.30455 | 72.13636 | 4 | 3 | 1 |
| ICE LEMON - TROPICAL MINT | 2 | 2020-02-01 | 1 | 7 | FALSE | FALSE | TRUE | ICE LEMON | 26.30455 | 72.13636 | 4 | 3 | 1 |

Fonte: Elaborada pelo autor (2020).

Em um primeiro momento, a Figura 38 demonstra que já temos previsões muito mais próximas dos valores reais do que utilizando as técnicas de Médias Móveis (Seção 4.2.1), Suavização Exponencial (Seção 4.2.2), ARIMA (Seção 4.2.3) e Redes Neurais Artificiais (Seção 4.3.1).

Figura 38 - Total de vendas por categoria para fevereiro/2020



Fonte: Elaborada pelo autor (2020).

A Figura 39 traz um gráfico de correlação entre os valores previstos e os encontrados, fazendo uma extensão das métricas de erros agora sob a ótica de produto, tratados individualmente.

Figura 39 – Correlação entre previsto (MLG) e realizado



Fonte: Elaborada pelo autor (2020).

Quanto mais próximos os pontos estiverem na linha diagonal no centro dos gráficos, maior a eficiência da previsão. O R^2 de 0.57 indica uma precisão moderada, pois como já citado, quanto mais próximo de 1, maior a acurácia do modelo. Podemos concluir que o RMSE e o MAE para os produtos estão bem abaixo das métricas encontradas para os produtos de forma agrupada, demonstrando que o modelo sugerido tem potencial para realizar previsões mais próximas da realidade. A Tabela 15 demonstra os resultados da previsão por categoria.

Tabela 15 – Previsão MLG por Categoria

| Categoria | Venda | Previsão | Erro |
|--------------------------|-------|----------|------|
| BLACK COFFEE | 28 | 25 | 11% |
| COFFEE MIX NESPRESSO | 287 | 243 | 15% |
| FROZEN | 6 | 3 | 50% |
| HOT COFFEE MIX NESPRESSO | 6 | 4 | 33% |
| HOT TRADICIONAIS | 7 | 4 | 43% |
| ICE LEMON | 1296 | 1508 | -16% |
| TRADICIONAIS | 821 | 881 | -7% |
| YAKULT | 610 | 688 | -13% |

Fonte: Elaborada pelo autor (2020).

Como em todas as técnicas de Aprendizagem de Máquina, a acurácia do modelo pode ter relação com a quantidade de dados disponíveis para realização do treinamento do modelo. Por conta da limitação de dados já citada em seções anteriores, surge a hipótese de que, com o passar do tempo, quanto mais dados este modelo receber como parâmetro de entrada, maior será sua acurácia.

4.2.7. Aplicação de Florestas Aleatórias

Para Lantz (2015), as Florestas Aleatórias combinam versatilidade e poder em uma única bordagem de Aprendizagem de Máquina e permite tratar conjunto de dados volumosos ao mesmo tempo em que oferece índices baixos de erros nas tarefas de aprendizagem se comparado com qualquer outro método.

Ainda segundo o autor, por conta da facilidade de uso, poder e versatilidade, as Florestas aleatórias estão se tornando rapidamente em um dos métodos de Aprendizagem de Máquina mais populares. Para Ramasubramanian & Singh (2019), a acurácia dos modelos baseados em Florestas Aleatórias tende a ser maior do que a maioria dos outros métodos baseados em árvores de decisão. A implementação neste estudo, assim como no caso do MLG, também foi realizada utilizando o pacote *caret* por meio da função *train()*. A Tabela 16 demonstra as previsões pelo algoritmo de Florestas Aleatórias.

Tabela 16 - Previsão para vendas de fevereiro de 2020

| Data | Vendas | Previsão | Margem de erro |
|------------|--------|----------|----------------|
| 01/02/2020 | 130 | 144 | -11% |
| 02/02/2020 | 125 | 125 | 0% |
| 03/02/2020 | 72 | 101 | -40% |
| 04/02/2020 | 100 | 115 | -15% |
| 05/02/2020 | 101 | 106 | -5% |
| 06/02/2020 | 100 | 118 | -18% |
| 07/02/2020 | 109 | 123 | -13% |
| 08/02/2020 | 194 | 155 | 20% |
| 09/02/2020 | 115 | 124 | -8% |
| 10/02/2020 | 73 | 105 | -44% |
| 11/02/2020 | 116 | 112 | 3% |
| 12/02/2020 | 110 | 119 | -8% |
| 13/02/2020 | 114 | 103 | 10% |
| 14/02/2020 | 109 | 120 | -10% |
| 15/02/2020 | 166 | 158 | 5% |
| 16/02/2020 | 115 | 122 | -6% |
| 17/02/2020 | 83 | 108 | -30% |
| 18/02/2020 | 95 | 100 | -5% |
| 19/02/2020 | 62 | 96 | -55% |
| 20/02/2020 | 88 | 106 | -20% |
| 21/02/2020 | 84 | 121 | -44% |
| 22/02/2020 | 122 | 150 | -23% |
| 23/02/2020 | 87 | 137 | -57% |
| 24/02/2020 | 119 | 116 | 3% |
| 25/02/2020 | 99 | 105 | -6% |
| 26/02/2020 | 93 | 115 | -24% |
| 27/02/2020 | 74 | 104 | -41% |
| 28/02/2020 | 80 | 117 | -46% |
| 29/02/2020 | 126 | 145 | -15% |

Fonte: Elaborada pelo autor (2020).

Tabela 17 - Métricas para erros de previsão utilizando Florestas Aleatórias

| Métrica | Valor |
|---------------------------------------|-------|
| Erro Absoluto Médio (MAE) | 18.58 |
| Erro Percentual Absoluto Médio (MAPE) | 20.17 |
| Raiz Quadrada do Erro Médio (RMSE) | 22.61 |

Fonte: Elaborada pelo autor (2020).

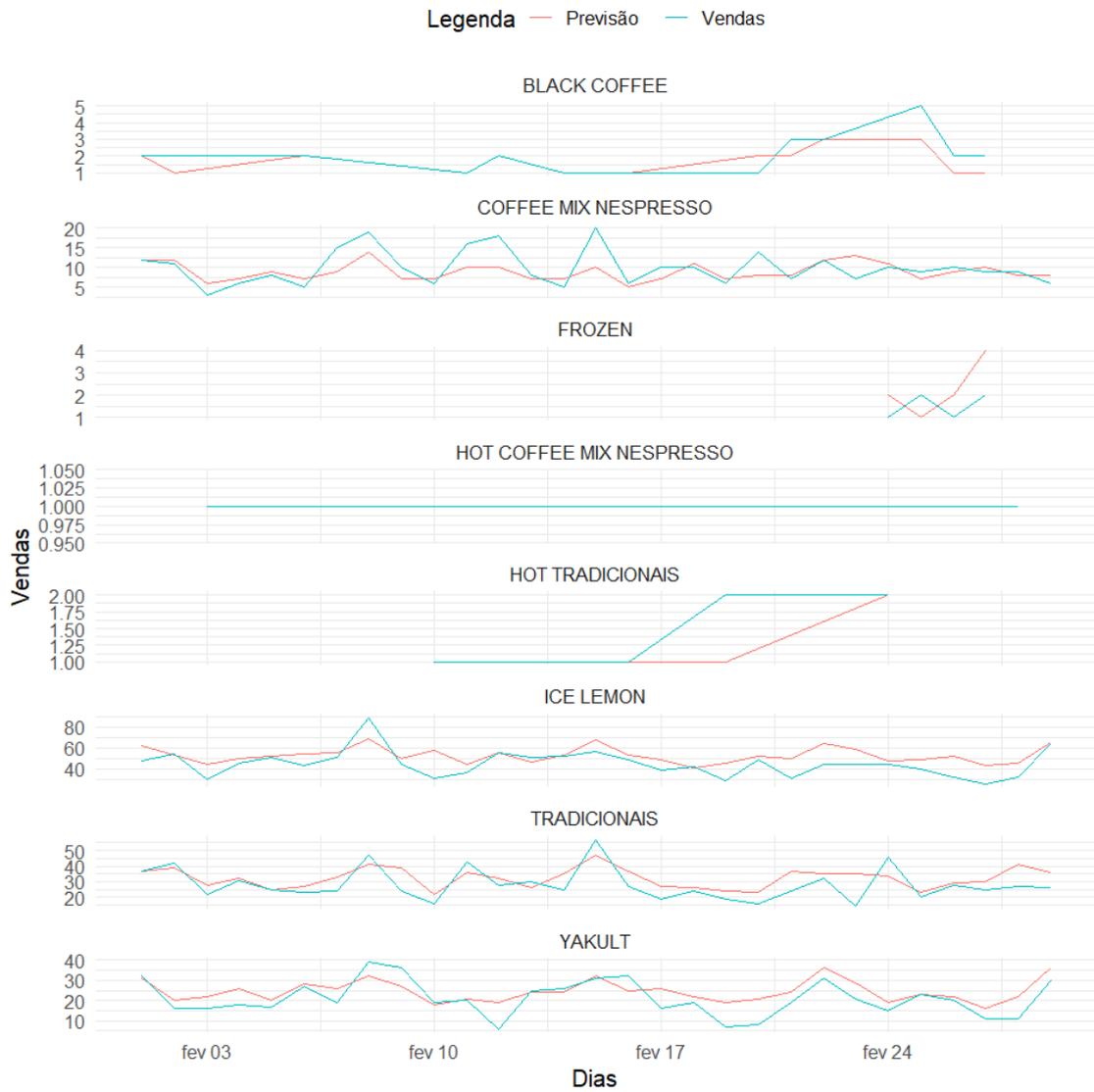
Assim como apresentado para o caso da previsão MLG, a Tabela 16 exhibe as métricas de erro para a previsão utilizando o algoritmo de Florestas Aleatórias.

Para esse modelo, podemos observar na Tabela 16 que todas as métricas (MAE, MAPE e RMSE) para o caso de Florestas Aleatórias, apresentam menores índices de acertos se comparado com os Modelos Lineares Generalizados.

Assim como no caso do MLG, o algoritmo criado para implementar os modelos de Aprendizagem de Máquina para este estudo, permitem que a função receba uma lista com os métodos de predição a serem testados e retorna então uma previsão detalhada por dia, por produto e por categoria de produto. Essa visão detalhada, permite uma melhor análise, se compararmos com a visão agrupada, que não faz distinção entre os produtos comercializados. A previsão detalhada permite uma visualização do conjunto de dados muito semelhante à apresentada na Figura 38, porém, trazendo as previsões relativas à utilização do algoritmo de Florestas Aleatórias. Podemos observar na Figura 40 a visualização da previsão, sendo a linha azul as vendas reais e a linha laranja a previsão.

Assim como no caso do MLG, o modelo de Florestas Aleatórias apresenta previsões mais próximas da realidade, como podemos ver no caso dos produtos da categoria “*Hot Coffee Mix Nespresso*”, onde as linhas estão sobrepostas o que indica que o modelo foi capaz de prever o que aconteceu de forma exata no conjunto de dados separados para teste, ou seja, as vendas reais do mês de fevereiro de 2020.

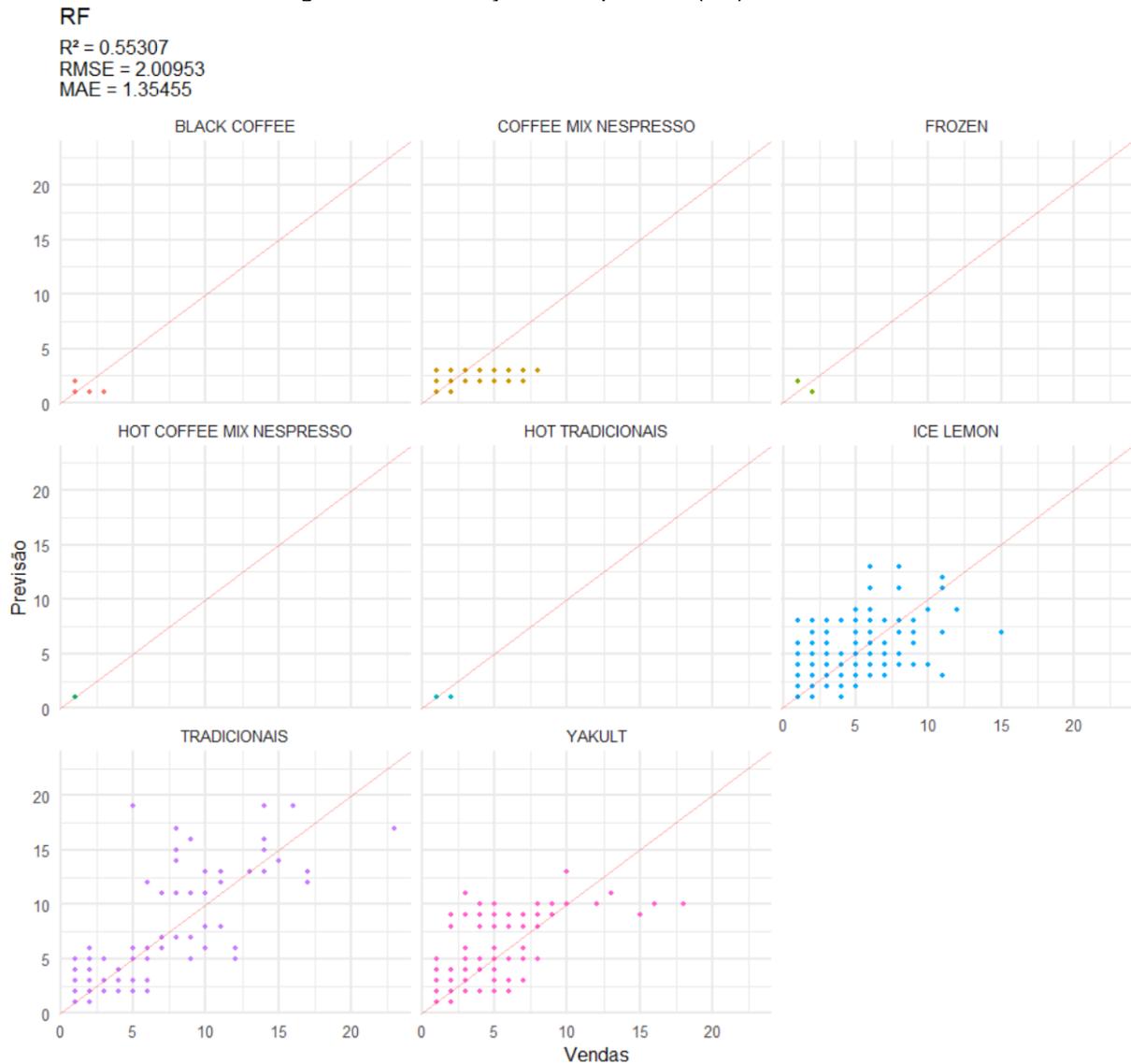
Figura 40 - Previsão de vendas para fevereiro/2020
 Total de vendas por categoria mês de Fevereiro/2020



Fonte: Elaborada pelo autor (2020).

A Figura 41 ilustra os gráficos de correlação entre a previsão e as vendas.

Figura 41 - Correlação entre previsto (RF) e realizado



Analisando a Figura 41, verificamos um R^2 de 0.55 o que indica uma previsão de qualidade moderada, e inferior ao valor de 0.57 encontrado no Modelo Linear Generalizado. Os valores do RMSE e do MAE também indicam que na comparação entre os dois modelos, por conta da disponibilidade de dados, o modelo MLG segue com erros menores do que o modelo de Florestas Aleatórias. A Tabela 17 exibe a margem de erro para as previsões por categoria.

Tabela 18 – Previsão de Florestas Aleatórias por Categoria de Bebidas

| Categoria | Venda | Previsão | Erro |
|--------------------------|-------|----------|------|
| BLACK COFFEE | 28 | 23 | 18% |
| COFFEE MIX NESPRESSO | 287 | 258 | 10% |
| FROZEN | 6 | 9 | -50% |
| HOT COFFEE MIX NESPRESSO | 6 | 6 | 0% |
| HOT TRADICIONAIS | 7 | 6 | 14% |
| ICE LEMON | 1296 | 1522 | -17% |
| TRADICIONAIS | 821 | 936 | -14% |
| YAKULT | 610 | 710 | -16% |

Fonte: Elaborada pelo autor (2020).

A Tabela 17 nos confirma que os erros produzidos pela previsão utilizando o algoritmo de Florestas Aleatórias encontra valores mais distantes da realidade, se comparado ao Modelo Linear Generalizado.

Na próxima seção, uma breve comparação entre os modelos será realizada para que a proposta do modelo preditivo para recomendação de estoque seja realizada.

4.3. COMPARATIVO ENTRE TÉCNICAS DE PREDIÇÃO APLICADAS

A aplicação de métodos de previsão progrediu substancialmente desde a década de 40 com os primeiros usos de Suavização Exponencial para prever a demanda e projetar o inventário de diversos itens em estaleiros da marinha norte americana. A introdução da metodologia Box-Jenkins e dos modelos ARIMA aproximaram a área acadêmica de campos de atuação que eram dominados por profissionais com experiência prática (MAKRIDAKIS; SPILLOTIS & ASSIMAKOPOULOS, 2018).

Para Tsoumakas (2018), a tarefa de previsão de vendas de alimentos é uma tarefa de previsão de séries temporais e abordagens estatísticas clássicas como ARIMA podem ser utilizadas para realização dessa tarefa. Porém, ainda segundo o autor, utilizar uma abordagem baseada em técnicas de aprendizagem de máquina para lidar com previsões de séries temporais é geralmente mais flexível e poderosa,

porque permite usar algoritmos modernos de aprendizagem, como por exemplo, modelos baseados em árvores, como é o caso de Florestas Aleatórias.

Tsoumakas (2018, p. 5) é categórico em afirmar que:

“Previsão de vendas para alimentos é uma tarefa importante para negócios modernos. Pode prover assistência em decisões de curto e longo prazo e pode levar a redução de custos e aumento de vendas.”⁶

Para a próxima etapa desta pesquisa, é importante analisar os resultados encontrados até o momento para definição do modelo preditivo mais adequado para responder ao problema identificado no início do presente trabalho. Percebemos pela Tabela 18, que com a evolução do experimento para a utilização de técnicas mais complexas, como o Aprendizado de Máquina, as métricas de erro apresentaram uma redução significativa, com exceção da Suavização Exponencial, que apresentou métricas com erros próximos a técnicas mais complexas.

Tabela 19 – Comparativo entre técnicas de predição

| Técnica | MAE | MAPE | RMSE |
|--------------------------------|-------|-------|-------|
| Médias Móveis | 51.75 | 56.33 | 55.20 |
| Suavização Exponencial | 20.68 | 21.43 | 27.75 |
| ARIMA | 33.65 | 35.55 | 39.64 |
| Redes Neurais | 20.10 | 18.12 | 28.34 |
| Modelos Lineares Generalizados | 14.51 | 14.68 | 20.05 |
| Florestas Aleatórias | 18.58 | 20.17 | 22.61 |

Fonte: Elaborada pelo autor (2020).

O modelo que se mostrou mais promissor, por conta dos índices de erros significativamente mais baixos que as demais, foram os Modelos Lineares Generalizados. Se comparado com outras técnicas de Aprendizagem de Máquina, como Redes Neurais e Florestas Aleatórias, demonstrou realizar previsões mais próximas da realidade, com MAE = 14.51, MAPE = 14.68 e RMSE = 20.05. Comparando os resultados de correlação levando em consideração o detalhamento da previsão, vemos que o MLG apresentou $R^2 = 0.57$ frente ao $R^2 = 0.55$ do método de Florestas Aleatórias.

⁶ Tradução livre a partir do trecho original: “Food sales prediction is an important task for the modern business. It can assist in both short and long term decision making and can lead to cost reduction and sales increase.”

Para demonstrar o consumo de recursos computacionais durante o processamento, a Tabela 19 exibe métricas relacionadas a esta visão comparando diferentes técnicas baseadas em Aprendizagem de Máquina. A máquina onde os testes foram executados possui um processador Intel i5 9600K 3.70 GHz, 16 GB com 2400 Mhz de Memória RAM e SSD M.2 NVMe de 500GB.

Tabela 20 – Recursos computacionais utilizados

| Técnica | Uso de Memória RAM | Tamanho do Arquivo | Duração | MAE | RMSE | R ² |
|--------------------------------|--------------------|--------------------|---------|---------|---------|----------------|
| Modelos Lineares Generalizados | 33.8 mb | 23.9 mb | 12,92s | 1.35651 | 1.98402 | 0.57243 |
| Florestas Aleatórias | 63.3 mb | 24.6 mb | 1h 33m | 1.35455 | 2.00953 | 0.55307 |

Fonte: Elaborada pelo autor (2020).

Analisando as métricas relacionadas a previsão, vemos que o MLG apresenta o $R^2 = 0.57243$, sendo o melhor entre os dois modelos comparados, demonstrando que a acurácia da previsão é mais alta. Além disso, a utilização de recursos computacionais é consideravelmente mais baixa, levando apenas aproximadamente 13 segundos para execução do algoritmo, bem como menor uso de memória e menor tamanho do arquivo gerado pela execução do modelo, quando comparado com a técnica de Florestas Aleatórias.

É possível concluir então, que com o conjunto de dados disponíveis – em que suas limitações já foram discutidas em seções anteriores – e por conta das métricas de acurácia da previsão serem mais elevadas e o consumo de recursos computacionais menor, que o melhor modelo para realizar a predição de estoque é a Técnica de Aprendizagem de Máquina baseada em Modelos Lineares Generalizados.

No Capítulo 5, o modelo será utilizado como ponto de partida para que uma recomendação de estoque seja realizada para a unidade da franquia Bubble Mix que foi objeto deste estudo.

5. O MODELO PREDITIVO DE ESTOQUE PROPOSTO

A Ficha Técnica de Preparo se demonstra útil para previsão do estoque, uma vez que descreve a quantidade de insumos utilizado na preparação de cada bebida oferecida no cardápio da unidade. A partir dos resultados da previsão de vendas da unidade em questão, foi necessário calcular a quantidade de cada ingrediente da FTP dos produtos.

A próxima etapa então, foi a criação de uma forma de armazenar as informações referentes aos ingredientes presentes na Ficha Técnica de Preparo para as bebidas, de modo que permitisse a mensuração de todos os insumos necessários para atender a demanda. A Figura 42 demonstra um recorte da FTP completa disponibilizada pela franqueadora, onde os ingredientes para cada bebida são apresentados.

Figura 42 - Recorte da FTP da bebida Yakult Cranberry

| 22 - YAKULT CRANBERRY | | |
|------------------------------|-------------------|----------------|
| Insumo | Quantidade | Unidade |
| Verde | 3,5 | gr. |
| Yakult | 1 | 1 |
| Açúcar | 15 | gr. |
| Essência | 25 | ml. |
| Copo | 1 | un. |
| Canudo | 1 | un. |
| Filme | 1 | un. |

Fonte: FTP da franqueadora Bubble Mix (2018).

A partir do recorte da FTP contendo os ingredientes, foi gerado um arquivo .CSV com as informações detalhando as receitas de todas as bebidas disponibilizadas no cardápio da franquia Bubble Mix. A Figura 43 demonstra um recorte do arquivo criado para posterior cálculo da quantidade de ingredientes necessários para atendimento a demanda prevista.

Figura 43 - Recorte do arquivo CSV contendo parte da FTP dos produtos

| Produto | Ingrediente | Quantidade | Unidade |
|---------------------------|------------------------------|------------|---------|
| YAKULT - CRANBERRY | Chá verde | 3.5 | g |
| YAKULT - CRANBERRY | Yakult | 1 | un |
| YAKULT - CRANBERRY | Açúcar | 15 | ml |
| YAKULT - CRANBERRY | Essência de cranberry | 25 | ml |
| YAKULT - CRANBERRY | Copo tradicional | 1 | un |
| YAKULT - CRANBERRY | Canudo | 1 | un |
| YAKULT - CRANBERRY | Filme | 1 | un |
| YAKULT - LARANJA VERMELHA | Chá verde | 3.5 | g |
| YAKULT - LARANJA VERMELHA | Yakult | 1 | un |
| YAKULT - LARANJA VERMELHA | Açúcar | 15 | ml |
| YAKULT - LARANJA VERMELHA | Essência de laranja vermelha | 25 | ml |
| YAKULT - LARANJA VERMELHA | Copo tradicional | 1 | un |
| YAKULT - LARANJA VERMELHA | Canudo | 1 | un |
| YAKULT - LARANJA VERMELHA | Filme | 1 | un |

Fonte: Elaborada pelo autor (2020).

O arquivo contendo o “livro” de receitas é importado pelo script criado no R Studio, criando um objeto que se torna disponível para consulta possibilitando assim que todas as vendas previstas para os produtos, determine a quantidade de ingredientes necessários para produzi-las. A Figura 44 exibe a sumarização realizada para saber exatamente quantas e quais bebidas foram previstas para serem comercializadas no horizonte contemplado.

Figura 44 - Recorte de produtos previstos sumarizados por categoria

| | | | |
|----|--------|---------------------------|-----|
| 56 | YAKULT | YAKULT - CRANBERRY | 41 |
| 57 | YAKULT | YAKULT - LARANJA VERMELHA | 70 |
| 58 | YAKULT | YAKULT - LICHIA | 26 |
| 59 | YAKULT | YAKULT - LIMÃO SICILIANO | 36 |
| 60 | YAKULT | YAKULT - MAÇÃ VERDE | 165 |
| 61 | YAKULT | YAKULT - MELÃO | 32 |
| 62 | YAKULT | YAKULT - MORANGO | 269 |
| 63 | YAKULT | YAKULT - NATURAL | 19 |
| 64 | YAKULT | YAKULT - PÊSSEGO | 52 |

Fonte: Elaborada pelo autor (2020).

Os dados gerados pela previsão são cruzados com os ingredientes listados no arquivo de receitas, para que seja calculada então a quantidade total de todos os insumos necessários para atendimento da demanda. A Figura 45 ilustra um recorte resultado deste cálculo.

Figura 45 - Cálculo de ingredientes para atender previsão

| | | | |
|----|----------------------|-------|----|
| 18 | Copo Frozen | 9 | un |
| 19 | Copo tradicional | 3187 | un |
| 20 | Creamer | 36150 | g |
| 21 | Essência de abacaxi | 800 | ml |
| 22 | Essência de abaxi | 1395 | ml |
| 23 | Essência de amaretto | 540 | ml |
| 24 | Essência de amêndoas | 480 | ml |
| 25 | Essência de avelã | 2220 | ml |
| 26 | Essência de baunilha | 6570 | ml |

Fonte: Elaborada pelo autor (2020).

É importante ressaltar que para que uma recomendação de estoque seja feita, é necessário realizar uma transformação dos insumos com suas quantias brutas, para o formato de embalagens comercializadas pelos fornecedores. Utilizando como exemplo o insumo “Copo Tradicional”, veremos que o modelo nos retorna à quantidade de 3.187 copos. Porém, o fornecedor que atende as franquias da Bubble Mix, comercializa os copos no formato de caixas com 1.000 unidades presentes em cada caixa. Para este caso em específico, sabemos então, que não podem ser comprados 187 unidades de copos avulsos, logo, o pedido deve ser arredondado para 4 caixas de 1.000 unidades, para que não corra o risco de faltar este item, que tem extrema importância na composição do produto.

Para tratar essas condições, a solução encontrada foi criar um arquivo CSV contendo as conversões das unidades de medida para o formato em que são comercializadas. A Figura 46 exibe um recorte do arquivo de conversões para cada um dos insumos presentes na FTP disponibilizada pela franqueadora:

Figura 46 - Conversões dos insumos

| | ingredient | package_divisor | package_label |
|----|--------------------------|-----------------|---------------|
| 1 | Açúcar | 10000 | baldes |
| 2 | Bolacha recheada oreo | 1 | pacotes |
| 3 | Calda de maracujá | 1000 | bisnagas |
| 4 | Canudo | 1000 | caixas |
| 5 | Cápsula espresso forte | 50 | caixas |
| 6 | Cápsula espresso leggero | 50 | caixas |
| 7 | Cápsula lungo forte | 50 | caixas |
| 8 | Cápsula lungo leggero | 50 | caixas |
| 9 | Chá de frutas vermelhas | 15 | caixas |
| 10 | Chá de hibisco | 1000 | pacote |
| 11 | Chá mate | 250 | caixinhas |
| 12 | Chá preto assam | 1000 | kg |
| 13 | Chá preto ceilão | 1000 | kg |
| 14 | Chá verde | 1000 | kg |
| 15 | Chantilly | 1000 | kg |
| 16 | Copo 110 ml | 50 | fardo |

Fonte: Elaborada pelo autor (2020).

Os insumos demonstrados na Figura 46, são então transformados para o formato de embalagem no qual são comercializados pelo fornecedor. Em ambos os casos, tanto quanto na lista de ingredientes, quanto no arquivo de transformação, foram projetados de forma flexível, para comportar mudanças. Para atualizar o formato de embalagem de um insumo, por exemplo, basta alterar o arquivo .CSV contendo os divisores e o formato do pacote. Uma mudança semelhante pode ser feita em uma receita, por exemplo, pois basta referenciar de qual bebida se trata, que o modelo automaticamente vai passar a computar a nova receita para aquele produto.

A implementação realizada na linguagem de programação R por meio da ferramenta R Studio, deixou o modelo flexível para comportar a inserção de novas bebidas em uma eventual troca de cardápio, bem como realizar previsões com horizontes pré-definidos. A previsão foi realizada para o mês de fevereiro de 2020, mas é possível fazer para períodos mais curtos ou mais longos, dependendo da avaliação e necessidade do gestor da unidade. O resultado do experimento permite a configuração das embalagens para compra de insumos junto a fornecedores. Essa flexibilização foi projetada para atender a eventuais mudanças possibilitando uma rápida adaptação para o contexto da unidade em que venha a ser utilizada.

Ao final do processo, a Figura 47 ilustra um recorte do que seria a recomendação de estoque para atendimento ao horizonte utilizado para realizar as predições. Neste caso, a lista de compras sugerida por esta pesquisa, seria o suficiente para atender a demanda prevista, diminuindo o desperdício e a imobilização do capital da unidade.

Figura 47 - Recorte da recomendação de estoque

| | ingredient | order_quantity | package_label |
|----|--------------------------|----------------|---------------|
| 1 | Açúcar | 8 | baldes |
| 2 | Bolacha recheada oreo | 4 | pacotes |
| 3 | Calda de maracujá | 3 | bisnagas |
| 4 | Canudo | 4 | caixas |
| 5 | Cápsula espresso forte | 1 | caixas |
| 6 | Cápsula espresso leggero | 1 | caixas |
| 7 | Cápsula lungo forte | 6 | caixas |
| 8 | Cápsula lungo leggero | 1 | caixas |
| 9 | Chá de frutas vermelhas | 34 | caixas |
| 10 | Chá de hibisco | 1 | pacote |
| 11 | Chá mate | 20 | caixinhas |
| 12 | Chá preto assam | 1 | kg |
| 13 | Chá preto ceilão | 3 | kg |
| 14 | Chá verde | 6 | kg |
| 15 | Chantilly | 1 | kg |
| 16 | Copo 110 ml | 1 | fardo |

Fonte: Elaborada pelo autor (2020).

No próximo capítulo serão apresentadas as Considerações Finais e perspectivas referentes ao presente estudo.

6. CONSIDERAÇÕES FINAIS

Esta pesquisa, de cunho aplicado, teve como tema a utilização de tecnologias para aprimoramento da gestão de estoques em uma franquia de *bubble tea*, considerando suas principais dificuldades no que diz respeito a gestão de estoques e explorando os motivos pelos quais um modelo preditivo de venda poderia contribuir para sanar essas questões.

Para contextualizar o problema de pesquisa investigado, foram apresentadas revisões bibliográficas sobre Gestão e Controle de Estoques e em seguida, foram identificadas Técnicas de Predição de Demandas que poderiam ser utilizadas no presente estudo. As previsões de vendas, realizadas de forma precisa e por meio de um método cientificamente comprovado, podem contribuir para uma gestão eficaz de estoques e influenciar diretamente na lucratividade em uma franquia de *bubble tea*.

A partir da identificação das técnicas, os dados disponibilizados pela empresa estudada, foram preparados e consolidados para que pudessem ser utilizados como dados de entrada para a aplicação das técnicas. Nesta etapa, uma Análise Exploratória dos dados disponíveis foi conduzida, com o objetivo de identificar variáveis de interesse para os modelos preditivos e aplicar técnicas qualitativas que contribuíssem à formação e consolidação do conjunto de dados para a etapa seguinte.

Os dados consolidados foram divididos entre os dados de treinamento dos modelos – período de 01 de setembro de 2019 até 31 de janeiro de 2020 - e os dados para “teste” dos resultados dos modelos, que compreendiam as vendas efetivadas no mês de fevereiro de 2020. Esses dados foram utilizados para testar os modelos sugeridos no presente estudo, com o objetivo de verificar sua acurácia por meio de avaliação de métricas. Após avaliação das técnicas de predição, o melhor modelo preditivo de acordo com as métricas de avaliação da acurácia e com o conjunto de dados criados, se mostrou o modelo baseado em Modelos Lineares Generalizados.

O modelo proposto foi utilizado para construir um conjunto de dados de previsão de vendas, e esse novo conjunto contendo as informações relativas a previsão de produtos separados por categoria, serviu de ponto de partida para identificação dos insumos necessários para atendimento da demanda prevista e uma

transformação desses insumos no formato de comercialização foi realizada para que a recomendação fosse então gerada. Como resultado, podemos concluir que a previsão de vendas pode contribuir positivamente para um nível de estoque suficiente para atendimento a demanda, sem que haja imobilização de capital ou incremento nos custos de manutenção do estoque.

Com base nos resultados da Análise Experimental conduzida na Seção 4.1, ficou evidente que a adoção de tecnologias e recursos computacionais, como no caso desta pesquisa a linguagem de programação R, contribuem para um entendimento mais profundo sobre os aspectos do negócio, sendo o resultado mais valioso a produção de novos conhecimentos, como a correlação entre as variáveis investigadas e a sua influência na flutuação de vendas na unidade em que o experimento foi conduzido. A aplicação de técnicas quantitativas e qualitativas demonstraram agregar valor e influenciar na lucratividade do negócio.

O experimento gerou resultados expressivos, especialmente na relação custo-benefício no que diz respeito à adoção de técnicas para gestão e controle de estoques aliadas a um modelo preditivo de vendas com recomendação de um estoque para atendimento a demanda esperada. De posse da recomendação, o gestor da unidade será capaz de investir o capital de giro na empresa apenas em insumos que sejam realmente necessários para atendimento a demanda. As Técnicas de Aprendizagem de Máquina se mostraram bastante promissoras, por se adaptarem ao contexto do negócio conforme a base de dados do ERP utilizado seja alimentada. Uma massa de dados maior pode prover mais informações e alimentar o modelo, que irá reagir de forma iterativa e incremental, melhorando seus resultados e conseqüentemente sua recomendação de estoque para atendimento a demanda.

É importante observar que os dados disponibilizados e suas limitações, apresentaram alguns obstáculos que geraram lacunas de conhecimento que não foram possíveis de serem contornadas. A ausência de um ciclo completo de vendas para análise, pode ter enfraquecido algumas conclusões, apesar de não inviabilizarem o experimento. É possível afirmar que com um maior acúmulo de dados no futuro, tanto a análise experimental quanto o experimento em si, possam ser replicados com resultados mais eficazes. O modelo construído permite a adaptação a este cenário

mais favorável, o que nos leva a concluir que poderá ser utilizado de maneira contínua, possibilitando novas descobertas “escondidas” nos dados.

A interdisciplinaridade desta pesquisa demonstrou por seus resultados que diferentes áreas de conhecimento como computação e gestão, quando aliadas, podem trazer benefícios tanto para o campo acadêmico quanto para a esfera profissional, pois a aplicação do conhecimento gerado em um problema do mundo real, contribui para a sustentabilidade dos negócios, trazendo não só um benefício de cunho científico, como também social.

Por fim, visando uma extensão das contribuições prestadas pelo presente estudo, surgem algumas oportunidades alvo de trabalhos futuros:

- Desenvolvimento de um modelo preditivo para temperatura e umidade do ar baseado em dados históricos;
- Desenvolvimento de mecanismos que implementam modelos de gestão de estoques aliados as previsões para atendimento a demanda;
- Automatização dos pedidos de compra baseado na recomendação e utilizando os prazos de entrega para garantir o contínuo abastecimento do estoque;
- Aprimoramento do modelo preditivo de vendas após aumento da massa de dados disponíveis para utilização.

REFERÊNCIAS

- ADEBANJO, D.; MANN, R. (2000). **Identifying problems in forecasting consumer demand in the fast-moving consumer goods sector**. *Benchmarking: An International Journal*, v. 7, p. 223-230, 2000.
- AHO, K.; DERRYBERRY, D.; PETERSON, T. (2014). **Model selection for ecologists: the worldviews of AIC and BIC**. *Ecology*, v. 95, p. 631–636, 2014.
- ALVES, E. J. (2013). **Métodos de bootstrap e aplicações em problemas biológicos**. Dissertação de Mestrado em Matemática — UNESP, p. 11-89, Rio Claro, 2013.
- ARMSTRONG, S. J. (2002). **Principles of forecasting**. Philadelphia: Kluwer Academic Publishers, 2002.
- ARUNRAJ, N. S.; AHRENS, D. (2015). **A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting**. *International Journal of Production Economics*, v. 170, p. 321–335, 2015.
- BALLOU, R. H. (2009). **Gerenciamento da cadeia de suprimentos: planejamento, organização e logística empresarial**. 5. ed. Porto Alegre: Bookman, 2009.
- BOROVYKH, A.; BOHTE, S.; OOSTERLEE, C, W. (2017). **Conditional time series forecasting with convolutional neural networks**. arXiv preprint arXiv:1703.04691, 2017.
- BOX, G. E. P.; JENKINS, G.M.; REINSEL, G. & LJUNG, G.M (2015). **Time Series Analysis: Forecasting and Control**. 5. ed. Hoboken: John Wiley & Sons, 2015.
- BOX, G.E.P.; JENKINS, G.M. (1976). **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1976.
- BREIMAN, L. (2001). **Random Forests**. *Machine Learning*, v. 45, p. 5–32, 2001.
- BREIMAN, L. (2001). **Statistical modeling: The two cultures**. *Statistical Science*, v. 16, p. 199–231, 2001.
- BUFFA, E. S. (1968). **Production inventory systems: planning and control**. 1. ed. Illinois: Richard D. Irwin, 1968.
- BUZBY, J.; WELLS, H.; AXTMAN, B.; MICKEY, J. (2009). **Supermarket loss estimates for fresh fruit, vegetables, meat, poultry, and seafood and their use in the ERS loss-adjusted food availability data**. *Economic Information Bulletin*. United States Department of Agriculture, v. 44, 2009.
- CHASE, C. (2013). **Demand-driven forecasting: a structured approach to forecasting**. 2. ed. Hoboken: John Wiley & Sons, 2013.
- CHNG, E. S; CHEN, S.; MULGREW, B. (1996). **Gradient radial basis function networks for nonlinear and nonstationary time series prediction**. *IEE Transactions on Neural Networks*, v. 7, p. 190-194. 1996.
- CHOPRA, S.; MEINDL, P. (2016). **Gestão da cadeia de suprimentos: estratégia, planejamento e operações**. 6. ed. São Paulo: Pearson Education do Brasil, 2016.

CLARK, M. (2013). **An Introduction to Machine Learning with Applications in R**. Center for Social Research, France: University of Notre Dame, 2013.

COELHO, L.C.; FOLLMANN, N.; RODRIGUEZ; TABOTA, C.M. (2008). **O efeito chicote na cadeia de abastecimentos**. Revista Mundo Logística. Logística e Supply chain management. v. 1, p. 20-25, 2008.

CORRÊA, H. L.; GIANESI, I. G. N.; CAON, M. (2007). **Planejamento, programação e controle da produção**. 5. ed. São Paulo: Atlas, 2007.

CRODA, R. M. C.; ROMERO, D. E. G.; MORALES, S. O. C. (2019). **Sales Prediction through Neural Networks for a Small Dataset**, International Journal of Interactive Multimedia and Artificial Intelligence, v. 5, p. 35-41, 2019.

CURITIBA. Secretaria do Estado da Educação. **Resolução N° 09/2019 – GS/SEED, de 1° de janeiro de 2019**. Disponível em: http://www.gestaoescolar.diaadia.pr.gov.br/arquivos/File/calendario/calendario_escolar_2019.pdf (Acessado em: 10 jun. 2020.)

DELL, M.; JONES, B.; OLKEN, B. (2013). **What Do We Learn from the Weather? The New Climate-Economy Literature**. Journal of Economic Literature, v. 52, p. 740-798, 2013.

DICKEY, D.A.; W.R. BELL; MILLER, R.B. (1986) **Unit roots in time series models: tests and implications**. American Statistician, v. 40, p. 12-26, 1986.

DIVAKAR, S.; RATCHFORD, B.T.; SHANKAR, V. (2005). **CHAN4CAST: a multichannel, multiregion sales forecasting model and decision support system for consumer-packaged goods**. Marketing Science, v. 24, p. 334–350, 2005.

DOGANIS, P.; ALEXANDRIDIS, A.; PATRINOS, P.; SARIMVEIS, H. (2006). **Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing**. Journal of Food Engineering, v. 75, p. 196–204, 2006.

DONSELAAR, V. K.; WOENSEL, T.; BROEKMEULEN, R.; FRANSOO, J. (2006). **Inventory control of perishables in supermarkets**. International Journal of Production Economics, v. 104, p. 462–472, 2006.

ERIKSSON, M. (2012). **Retail Food Wastage a Case Study Approach to Quantities and Causes**. Swedish University of Agricultural Sciences, Uppsala, 2012.

FERNANDES; E. M. G.P (1999). **Estatística aplicada**, Universidade do Minho: Braga, 1999.

FOWDUR, T. P.; BEEHARRY, Y.; HURBUNGS, V.; BASSOO, V.; RAMNARAIN-SEETOHUL, V. **Internet of Things and Big Data Analytics Toward Next-Generation Intelligence**. Mauritius: Springer International Publish, 2018.

FOZ DO IGUAÇU. Prefeitura Municipal. **Decreto Municipal N° 26.945, de 8 de janeiro de 2019**. Disponível em: <http://www.pmfi.pr.gov.br/ArquivosDB?idMidia=107142> (Acessado em: 10 jun. 2020).

FOZ DO IGUAÇU. Prefeitura Municipal. **Decreto Municipal N° 27.806, de 6 de janeiro de 2020**. Disponível em:

https://www5.pmf.br/admin_foz/diarios_oficiais//163.pdf (Acessado em: 10 jun. 2020.)

GHATAK, A. (2017). **Machine Learning with R**. 1. ed., India: Springer, 2013.

GIANESI, G. N.; BIAZZI, L. J. (2011). **Gestão estratégica dos estoques**. Revista de Administração da Universidade de São Paulo, v. 46, p. 290-304, 2011.

GONÇALVES, P.S. (2010). **Administração de Materiais**. 3. ed. Rio de Janeiro: Editora Campus, 2010.

GRANGER, C. W. J. (1978). **Seasonality: Causality, Interpretation, and Implications, in Seasonal Analysis of Economic Time Series**. Economic research report. U.S. Department of Commerce, Bureau of the Census, p. 33-46, 1978.

GUSTAVSEN, G. (2005). **Public policies and the demand for carbonated soft drinks: a censored quantile regression approach**. XI Congress of the European Association of Agricultural Economists, The Future of Rural Europe in the Global Agri-Food System, 2005.

GUTIERREZ, D (2015). Machine Learning and Data Science: **An Introduction to Statistical Learning Methods with R**. 1. ed. Technics Publications, LLC. Basking Ridge, NJ, EUA, 2015.

HANKE, J. E.; WICHERN, D. (2014). **Business Forecasting**. 9. ed. Harlow: Pearson Education Limited, 2014.

HARRIS, F. W. (1913). **How many parts to make at once**. Factory, The Magazine of Management, v. 10, p. 135-136, 1913.

HAYES, R.; UPTON, D. (1998). **Operations based strategy**. California Management Review, v. 40, v. 4, p. 8-25, 1998.

HOPFIELD, J. J. (1982). **Neural Networks and physical systems with emergent collective computational abilities**. Proceedings of the National Academy of the Sciences of the USA. v. 79, p. 2554-2558, 1982.

HOSHMAND, R. A. (2010). **Business Forecasting: A Practical Approach**. 2. ed. New York: Routledge, 2010.

HYNDMAN, R.; ATHANASOPOULOS, G.; BERGMEIR, C.; CACERES, G.; CHHAY, L.; O'HARA-WILD, M.; PETROPOULOS, F.; RAZBASH, S.; WANG, E.; YASMEEN, F. (2020). **forecast: Forecasting functions for time series and linear models**. R package version 8.12, <Disponível em: <http://pkg.robjhyndman.com/forecast>>.

HYNDMAN, R.J.; ATHANASOPOULOS, G. (2018). **Forecasting: principles and practice**, 2. edition, OTexts: Melbourne, Australia, 2018.

KOHONEN, T. (1982). **Self-organized formation of topologically correct feature maps**. Biologic Cybernetics, v. 43, p. 59-69, 1982.

KOLASSA, S.; SIEMESSEN, E. (2016). **Demand Forecasting for Managers**. New York: Business Expert Press, 2016.

KUHN, M. (2020). **caret: Classification and Regression Training**. R package version 6.0-86. <Disponível em: <https://CRAN.R-project.org/package=caret>>

- LAKATOS, E.; MARCONI, M. A (1992). **Metodologia do trabalho científico**. 4. ed, São Paulo: Atlas, 1992.
- LANCASTER, G. A.; LOMAS, R. A. (1985). **Forecasting for sales and materials management**. Londres: Macmillan Publishers, 1985.
- LANTZ, B. (2015). **Machine Learning With R: Discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R**. PACKT Publishing, Birminham, 2015.
- LIAW, A.; WIENER, M. (2002). **Classification and Regression by randomForest**. R News, v. 2/3, 2002.
- LIN, G. F.; WU, M. C. (2011). **An RBF network with a two-step learning algorithm for developing a reservoir inflow forecasting model**. Journal of Hydrology, v. 405, p. 439–450, 2011.
- LOURENÇÃO, A. M.; BAPTISTA, E. C.; SOLER, E. M.; SOUZA, F. B.; CHERRI, A. C. (2017). **Mixed-Integer Nonlinear Model for Multiproduct Inventory Systems with Interior Point and Branch and Bound Method**. IEEE Latin America Transactions, v. 15, p. 744-752, 2017.
- MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. (2018). **Statistical and Machine Learning forecasting methods: Concerns and ways forward**. PLoS ONE, v. 13, p. 1-26, 2018.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. (1997). **Forecasting: Methods and Applications**. 3. ed., New York: John Wiley & Sons, 1997.
- MARTINS, P. G; ALT, P. R. C. (2005). **Administração de materiais e recursos patrimoniais**. 7. ed. São Paulo: Saraiva, 2005.
- MIRANDA, F. A.; FREITAS, S. R. C.; FAGGION, P. L. (2009). **Integração e interpolação de dados de anomalias ar livre utilizando-se a técnica de RNA e krigagem**. Boletim de Ciências Geodésicas, v. 15, p. 428-443, 2009.
- MOON, A. M. (2013). **Demand and Supply Integration: A Key to World-Class Demand Forecasting**. New Jersey: FT Press, 2013.
- NELDER, J.; WEDDERBURN, R. (1972). **Generalized Linear Models**. Journal of the Royal Statistical Society, v. 135, p. 370-384, 1972.
- NOVAES, A. G. (2007). **Logística e gerenciamento da cadeia de distribuição**. 3. ed. Rio de Janeiro: Elsevier, 2007.
- PAULA, G. A. (2013). **Modelos de regressão com apoio computacional**. Instituto de Matemática e Estatística Universidade de São Paulo, 2013.
- PEARCE, C. (1974). **Prediction Techniques for Marketing Planners**. Journal of Marketing Research. v. 24, 1974.
- POZO, H. (2002). **Administração de Recursos Materiais e Patrimoniais – Uma Abordagem Logística**. 2. ed. São Paulo: Atlas, 2002.

- R CORE TEAM (2015). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- RAMANATHAN, U.; MUYLDERMANS, L. (2010). **Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK**. *International Journal of Production Economics*, v. 128, p 538–545, 2010.
- RAMASUBRAMANIAN, K.; SINGH, A. (2019). **Machine Learning Using R: With Time Series and Industry-Based Use Cases in R**. 2. Ed. Springer Science+Business, New York, NY, 2019.
- RICHARDS, G. (2014). **Warehouse management: a complete guide to improving efficiency and minimizing costs in the modern warehouse**. 2. ed., Hong Kong: Kogan Page, 2014.
- ROGERS, P.; RIBEIRO, K.; ROGERS, D. (2004). **Avaliando o risco na gestão financeira de estoques**. *Anais do VII Simpósio de Administração da Produção, Logística e Operações Internacionais – SIMPOI – FGV-EAESP*, São Paulo, 2004.
- ROSA, H.; MAYERLE, S. F.; GONÇALVES, M. B. (2010). **Controle de estoque por revisão contínua e revisão periódica: uma análise comparativa utilizando simulação**. *Revista de Produção*, v. 20, p. 626-638, 2010.
- ROSS, D. F. (2015). **Distribution Planning and Control Managing in the Era of Supply Chain Management**. 3. ed. Chicago: Springer, 2015.
- SCHERR, F.C. (1989). **Modern Working Capital Management**. 1. ed. New York: Prentice-Hall, 1989.
- SETHI, S.; YAN, H.; ZHANG, H. (2005). **Inventory and Supply Chain Management with Forecast Updates**. *International Series in Operations Research & Management Science*. New York: Springer-Verlag, 2005.
- SHIN, Y; GHOSH, J. (1995). **Ridge polynomial networks**. *IEE Transactions on Neural Networks*, v. 3, p. 610-622. 1995.
- SHUKUR, O. B., & LEE, M. H. (2015). **Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA**. *Renewable Energy*, v. 76, p. 637–647, 2015.
- SICKLE, J. J. V.; BEILOCK, R. (1981). **An Interactive Forecasting Technique for Perishable Products**. In *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, p. 184–212. Ames, USA, 1981.
- SILVA, L. A.; PERES, M. S.; BOSCARIOLI, C. (2016). **Introdução à Mineração de Dados**. 1. ed. Rio de Janeiro: Elsevier, 2016.
- ŠTULEC, I.; PETLJAK, K.; NALETINA, D. (2019). **Weather impact on retail sales: How can weather derivatives help with adverse weather deviations?** *Journal of Retailing and Consumer Services*, v. 49, p 1–10, 2019.
- TRAPLETTI, A.; HORNIK, K. (2019). **tseries: Time Series Analysis and Computational Finance**. R package version 0.10-47. <Disponível em: <https://cran.r-project.org/web/packages/tseries/index.html>>

- TSOUMAKAS, G. (2018). **A survey of machine learning techniques for food sales prediction**. Artificial Intelligence Review, v. 52, p. 441-447, 2018.
- TURKMAN, M. A. A; SILVA, G. L. (2000). **Modelos Lineares Generalizados - da teoria à prática**, VIII Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa, 2000.
- VASCONCELLOS, F; CAVALCANTI E.; BARBOSA L. (2002). **Menu: como montar um cardápio eficiente**. São Paulo: Roca; 2002.
- VERIKAS, A.; VAICIUKYNAS, E.; GELZINIS, A.; PARKER, J.; OLSSON, M. C. (2016). **Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness**. Sensors 2016, p. 1-26, 2016.
- VIANA, J. J. (2010). **Administração de Materiais: Um enfoque prático**. 1. ed. São Paulo: Atlas, 2010.
- VORST, J. V. D.; BEULENS, A.; WIT, W. D.; BEEK, P. V. (1998). **Supply chain management in food chains: Improving performance by reducing uncertainty**. International Transactions in Operational Research, v. 5, p. 487–499, 1998.
- VU, T. B. (2015). **Seeing the Future: How to Build Basic Forecasting Models**. New York: Business Expert Press, 2015.
- WERNER, L.; RIBEIRO, J. L. (2003). **Previsão de Demanda: Uma Aplicação dos Modelos Box-Jenkins na Área de Assistência Técnica de Computadores Pessoais**. Gestão & Produção, v. 10, p. 476-67, 2003.
- ZHANG, G.; PATUWO, B. E; HU, Y. M. (1998). **Forecasting with artificial neural networks: The state of the art**. International Journal of Forecasting, v. 14, p. 35–62, 1998.
- ZHANG, Q.; BENVENISTE, A. (1992). **Wavelet networks**. IEEE Transactions on Neural Networks, v. 3, p. 889-898, 1992.

ANEXO

Bubble Asia Comercio de Chas Ltda
Av. Presidente Getúlio Vargas, 3620, Sala 603 - Água Verde
Curitiba, PR, Brasil.
contato@bubblemixtea.com
www.bubblemixtea.com

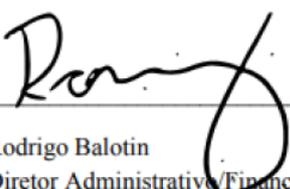


AUTORIZAÇÃO PARA REALIZAÇÃO DE PESQUISA PARA MESTRADO PROFISSIONAL

Informamos para os devidos fins, que autorizamos o Sr. Rodrigo Pereira Fontes, inscrito sob o CPF 058.476.439-18, R.G 7.016.334-9, aluno regular do Mestrado PPGTGS: Tecnologias, Gestão e Sustentabilidade da Universidade Estadual do Oeste do Paraná (UNIOESTE), a realizar a pesquisa intitulada: "PROPOSTA DE MODELO PREDITIVO DE VENDAS PARA UMA FRANQUIA DE BUBBLE TEA PARA RECOMENDAÇÃO DE ESTOQUE", respeitando as seguintes limitações:

- Omitir dados específicos de unidades que não seja de propriedade do aluno citado nesta autorização;
- Utilizar os dados de vendas da franquia Cataratas JL Shopping, do período de setembro de 2019 a fevereiro de 2020;
- Omitir dados relacionados a parceiros, fornecedores e demais ferramentas técnicas;
- Respeitar o acordo de confidencialidade descrito no contrato de franquia firmado entre as partes.

Destacamos que a Bubble Mix Tea é comprometida com esforços empreendidos para estreitamento de laços entre universidades e empresas, gerando conhecimento e beneficiando a todos os interessados na discussão de temas relevantes para o ecossistema empresarial.



Rodrigo Balotin
Diretor Administrativo/Financeiro
Bubble Mix Franqueadora