UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE TECNOLOGIA
CURSO DE ENGENHARIA DE PRODUÇÃO

ANA CAMILLA COÊLHO DE MACÊDO

COMPARANDO MODELOS CLÁSSICOS DE SÉRIES TEMPORAIS E APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DE DEMANDA NA INDÚSTRIA DE BEBIDAS

Caruaru
2022

ANA CAMILLA COÊLHO DE MACÊDO

COMPARANDO MODELOS CLÁSSICOS DE SÉRIES TEMPORAIS E APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DE DEMANDA NA INDÚSTRIA DE BEBIDAS

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção

Orientador: Prof. Dr. Caio Bezerra Souto Maior

Caruaru

2022

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Macêdo, Ana Camilla Coêlho de.
Comparando modelos clássicos de séries temporais e aprendizagem de máquina para previsão de demanda na indústria de bebidas / Ana Camilla Coêlho de Macêdo. - Caruaru, 2022.
54 : il., tab.

Orientador(a): Caio Bezerra Souto Maior
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, Engenharia de Produção - Bacharelado, 2022.
Inclui referências, apêndices.

1. Séries Temporais. 2. Previsão de Demanda. 3. ARIMA. 4. Support Vector Machine. 5. Holt-Winters. I. Maior, Caio Bezerra Souto. (Orientação). II. Título.

670 CDD (22.ed.)

ANA CAMILLA COÊLHO DE MACÊDO

COMPARANDO MODELOS CLÁSSICOS DE SÉRIES TEMPORAIS E APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DE DEMANDA NA INDÚSTRIA DE BEBIDAS

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção

Aprovada em: 19/05/2022.

BANCA EXAMINADORA

_____

Prof. Dr. Caio Bezerra Souto Maior (Orientador)
Núcleo de Tecnologia - UFPE

_____

Prof. Dr. Luciano Carlos Azevedo da Costa (Examinador Interno)
Núcleo de Tecnologia - UFPE

_____

Prof. Dr. Thyago Celso Cavalcante Nepomuceno (Examinador Interno)
Núcleo de Tecnologia - UFPE

AGRADECIMENTOS

Inicialmente, agradeço a Deus por ter me fortalecido todos os dias e, com a interseção de Nossa Senhora, me guiado durante todo esse percurso.

Agradeço à Painho (in memoriam), por todo amor e apoio dados desde sempre e até o fim. Sei que de onde estás, o senhor está me abençoando e orgulhoso de mim. À Mainha, meu lugar de segurança e de amor, por toda força durante o tempo de graduação e por toda confiança. Vocês são os amores da minha vida! Obrigada por todo incentivo e oportunidades que vocês me deram.

Aos meus irmãos, por todo apoio, preocupação e carinho durante todo esse tempo, e por terem sido presença quando tudo parecia desabar. Em especial, agradeço à minha irmã, Cibelle, que é minha grande inspiração desde criança, pela escuta, pelo companheirismo e por ser tão presente mesmo estando longe. Essa conquista também é de vocês!

Aos meus sobrinhos, que são fonte de alegria e leveza na minha vida. A busca por ser alguém melhor para vocês me faz ser alguém melhor para o mundo.

Aos amigos de longa data, que estiveram comigo durante todo esse processo e para além dele. Vocês foram casa, afeto, apoio e um escape nos momentos mais difíceis.

Aos amigos que a universidade me trouxe. Vocês fizeram desse caminho um lugar menos árduo e mais afetuoso. Foram, muitas vezes, a motivação que eu precisava para continuar. Todos os momentos que tivemos – desde os mais felizes até os de maior tensão – serão sempre lembrados.

Aos professores que tive durante a graduação e que me inspiraram a ser uma profissional melhor. Em especial, agradeço ao meu orientador, Prof. Caio, por toda disponibilidade, suporte e confiança. Com certeza, o seu auxílio foi parte fundamental para a realização deste trabalho.

Ao meu gestor de estágio e ao meu analista, pela disponibilidade e por terem acreditado desde o início na proposta deste trabalho.

À UFPE – CAA, lugar que me ensinou tanto sobre a vida profissional e que, acima de tudo, me ensinou sobre o mundo, educação e luta. Sou orgulhosa de ter feito parte desse lugar e de ser fruto da interiorização das universidades públicas.

Por fim, agradeço a todos que, de forma direta ou indireta, contribuíram para que este trabalho fosse realizado e para que eu pudesse chegar até aqui.

“Um passo à frente e você não está mais no mesmo lugar.”

Chico Science & Nação Zumbi (1996)

RESUMO

Devido ao crescimento da competitividade no mercado, a previsão de demanda tornou-se uma ferramenta fundamental como forma de gerenciamento da produção e identificação de novas oportunidades para a empresa. A alta demanda e mudança nas necessidades do consumidor fazem com que o setor de bebidas setor esteja cada vez mais empenhado na busca por uma maior eficiência e pela satisfação do cliente. Neste contexto, um dos objetivos da análise de séries temporais é o de realizar previsões a partir de dados históricos e, a partir disso, ter a possibilidade de tomar as decisões necessárias com maior segurança. Com o início da pandemia de COVID-19, o mercado sofreu inúmeras mudanças e as necessidades do consumidor mudaram, afetando diretamente as vendas de bebidas. Neste trabalho, modelos clássicos de séries temporais – Holt-Winters e ARIMA – e de aprendizagem de máquina – *support vector machine* – foram utilizados para realizar previsões de demanda a partir de séries de dados históricos do volume de bebidas de um centro de distribuição direta de bebidas. Os dados utilizados são de janeiro de 2018 até dezembro de 2021 e foram estratificados em nove séries – volume total de bebidas vendido pela operação, por tipo de bebida (cerveja e NAB) e em seis canais de vendas. Os modelos foram comparados a partir das previsões realizadas para até 15 passos (meses) à frente utilizando as métricas de erro RMSE e MAPE, apresentando assim dois modelos com melhores desempenhos: ARIMA e SVM. Ao final, constatou-se que não há um modelo que seja o mais indicado para ser usado em todas as séries – os três modelos obtiveram bom desempenho em pelo menos uma delas –, visto que suas performances variaram de acordo com a série analisada.

Palavras-chave: Séries Temporais. Previsão de Demanda. Indústria de bebidas. Holt-Winters. ARIMA. *Support Vector Machine*.

ABSTRACT

Due to the growth of competitiveness in the Market, demand forecasting has become a fundamental tool to manage production and identify new opportunities for the company. In the beverage industry is no different, the high demand and change in consumer needs make this sector increasingly committed to the pursuit of greater efficiency and customer satisfaction. One of the objectives of time series analysis is to make predictions from historical data and, from this, have the possibility of make the necessary decisions more securely. With the onset of the COVID-19 pandemic, the market has undergone numerous changes and consumer needs changed, directly affecting beverage sales. In this work, classic models of time series – Holt-Winters and ARIMA – and machine learning – Support vector machine – were used to perform demand forecasts from several historical data series of beverage volume from a beverage direct distribution center. The data used are from January 2018 to December 2021 and were stratified into nine series – total volume of beverages sold by the operation, by type of beverage (beer and NAB) and in six sales ways. The models were compared based on the predictions made for up to 15 steps (months) ahead using the RMSE and MAPE error metrics, revealing two models with higher performances: ARIMA and SVM. By the end, it was found that there is not a model that is the most suitable to be used in all series – the three models performed well in at least one of them –, since their performances varied according to the series analyzed.

Keywords: Time Series. Demand Forecast. Beverage Industry. Holt-Winters. ARIMA. Support Vector Machine.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O setor de bebidas, bem como o mercado empresarial, busca manter-se dinâmico em um ambiente altamente competitivo e de grandes mudanças no comportamento do consumidor, onde predominam as empresas que estão mais preparadas e organizadas. Dados do IBGE (2018) estimam que as indústrias de bebidas somam aproximadamente 1,2% de todas as indústrias de transformação no Brasil – tendo as fábricas de bebidas alcoólicas como maior participação – e 1,8% no nordeste – que tem maior participação com fábricas de bebidas não alcoólicas (POLARY, 2021). Além disso, em 2019 tal indústria obteve um faturamento equivalente a 1,9% do PIB do país (ABIA, 2020).

De acordo com JUNIOR (2017) e JUNIOR et al. (2014), refrigerantes e cervejas são os principais produtos do setor de bebidas e com o maior valor de vendas, sendo mais de 75% do valor total e 82% do volume total produzido. Por apresentar altas temperaturas e o clima tropical durante todo o ano em grande parte das regiões, o Brasil conta com um cenário ideal para o consumo de bebidas geladas, sendo o terceiro lugar no mundo com maior consumo e produção de bebidas (JUNIOR, 2017).

A indústria brasileira de bebidas tem se reinventado a partir de novas tecnologias e produtos, a fim de alcançar um mercado de consumidores de classes variadas. Com o início da pandemia de COVID-19, o cenário de consumo de bebidas foi totalmente modificado impactando também o comportamento do consumidor, que acaba sendo refletido nas vendas (VIANA, 2020, 2021).

Com o fechamento de estabelecimentos para consumo devido às medidas de isolamento durante a pandemia, é possível que a indústria de bebidas se remodele para atender as novas necessidades dos consumidores, com oportunidade para a criação de novos modelos de negócio. Essas mudanças de comportamento mostram uma tendência de crescimento no consumo de bebidas (alcoólicas ou não) nas residências, de forma a impactar diretamente o canal *"on trade"* que são os lugares onde as bebidas são consumidas (ex. bares e restaurantes). É possível também observar que uma das mudanças do padrão de consumo é a consolidação dos aplicativos de delivery para compra e venda de bebidas, sendo uma das oportunidades aproveitadas por grandes empresas.

Com base nesse ambiente dinâmico, as decisões quanto ao planejamento para os direcionamentos futuros são de extrema importância (LEMOS, 2006). A previsão de demanda se mostra como um elemento importante para a tomada de decisão, auxiliando a empresa a ter sucesso em seu planejamento estratégico e a melhorar sua eficiência (MAKRIDAKIS, 1996;

MONTGOMERY; JOHNSON; GARDINER, 1990). De fato, a previsão de demanda serve como um guia para a política de decisões, sejam elas de curto, médio ou longo prazo. Os diferentes horizontes de tempo ajudam a empresa na identificação de situações que necessitem de importantes decisões, comparando as ações tomadas com a direção prevista (LEMOS, 2006).

Sendo a análise de séries temporais um método quantitativo de previsão, TUBINO (2009) destaca que as previsões baseadas nelas partem do conceito de que a estimativa futura de uma variável será uma projeção apenas das suas observações passadas, de forma que não sofre nenhuma influência de outras variáveis. Os efeitos causados por previsões ruins podem afetar desde uso ineficiente dos recursos até o estoque, fazendo com que o nível de serviço da empresa seja diretamente afetado, sendo sentido pelo consumidor final (NENNI; GIUSTINIANO; PIROLO, 2013).

Dada a importância do tema, vários trabalhos vêm sendo realizados abordando os mais diversos cenários. BARZOLA-MONTESES et al. (2019) apresentam um modelo preditivo de séries temporais para o sistema de produção de energia hidrelétrica no Equador a partir da comparação dos modelos ARIMA e ARIMAX, mostrando o ARIMAX como o modelo mais indicado para o problema. ŠENKOVÁ et al. (2021) mostram uma análise dos impactos da pandemia de COVID-19 no turismo termal na Eslováquia, e utilizam o modelo ARIMA para a modelagem e previsão da capacidade e do desempenho das instalações de SPAs a partir de uma série temporal. Cong et al. (2019) utilizam o modelo SARIMA para prever os casos de gripe em Mainland, China, a partir de dados mensais. O modelo que melhor se ajustou à variação sazonal dos dados e que obteve maior desempenho quanto a incidência de casos foi o $(1,0,0)(0,1,1)[12]$.

MAIOR et al. (2016) comparam o uso da metodologia *empirical mode decomposition* (EMD) + *support vector machine* (SVM) com o uso somente do SVM para dados relacionados a confiabilidade com o intuito de estimar o tempo de vida útil remanescente (RUL) em dois casos diferentes: para turbocompressores em um motor a diesel e para motor a diesel de submarino. Para as duas situações, os resultados indicam o modelo EMD + SVM como o mais satisfatório por ter os resultados superiores; Já RADHIKA & SHASHI (2009) comparam o uso do SVM com o modelo *multilayer perceptron* (MLP), utilizando dados meteorológicos da Universidade de Cambridge para prever a temperatura máxima de um dia baseada nas temperaturas máximas dos dias anteriores. Ao final, os resultados mostram que o desempenho do SVM é superior ao MLP em todos os testes realizados.

CORTEZ & DONATE (2014) utilizam seis séries de dados, de diversas situações (ex. demanda de gasolina, demanda de passageiros, número de nascimentos diários), para comparar

as performances dos modelos *Holt-Winters* (HW), ARIMA, *evolutionary artificial neural network* (EANN) e *evolutionary* SVM (ESVM) – que foi dividido em *global* ESVM (GESVM) e *decomposition* ESVM (DESVM). Ao analisar os resultados de todos os modelos utilizados – a partir das métricas %RSE e %SMAPE –, o DESVM se destaca como o melhor para ambos os casos em três das seis séries analisadas, seguido do HW, ARIMA, EANN e GESVM que se destacam em apenas uma métrica de uma das séries.

FANIBAND & SHAAHID (2021) comparam os modelos *k-nearest neighbors* (kNN), *randon forest* (RF), *support vector regression* (SVR), HW e ARIMA, em uma série de tempo univariada para previsão a curto prazo da velocidade do vento como forma de aumentar a eficiência da geração de energia em sistemas eólicos. Ao final, os resultados indicam que os modelos RF, kNN e SVR – modelos baseados em aprendizagem de máquinas – apresentam um desempenho superior ao HW e ao ARIMA.

Apesar dos diferentes cenários encontrados na literatura sobre previsão a partir de séries temporais – e das inúmeras metodologias utilizadas –, poucos são os trabalhos encontrados que fossem relacionados à indústria de bebidas e comparem modelos clássicos e de aprendizado de máquinas, principalmente relacionados ao cenário que vivemos atualmente. Dessa forma, é válido dizer que no cenário atual, a utilização de modelos de previsão em séries temporais para a indústria de bebidas traz uma nova forma de enxergar as oportunidades e dar um pouco mais de segurança para tomadas de decisão.

1.1 OBJETIVOS

1.1.1 Geral

Apresentar e comparar modelos de previsão de demanda para um Centro de Distribuição Direta de bebidas localizado na cidade de Caruaru/PE, a fim de encontrar qual o modelo mais indicado para as situações analisadas.

1.1.2 Específicos

- Fazer revisão da literatura;
- Se familiarizar com a linguagem computacional para utilização dos modelos;
- Coletar dados;
- Aplicar modelos clássicos e de *aprendizado de máquina* nas séries temporais;
- Comparar os modelos clássicos de previsão utilizados nas séries.

1.2 JUSTIFICATIVA

Em um mercado consumidor que vive em constante transformação, a previsão de demanda possibilita que os gestores consigam prever o futuro de forma que possam planejar ações que melhor se adaptem ao que pode acontecer (KRAJEWSKI; RITZMAN; MALHOTRA, 2009; TUBINO, 2009). Desta forma, a previsão passa a ser um direcionar de negócios e se torna a base do planejamento estratégico – tanto na produção, quanto nas vendas e financeiro.

Apesar disso, é importante lembrar que as previsões apresentam erros, então, além de um modelo de previsão que os minimize ao máximo, contar com a experiência do gestor também é de grande importância, ajudando ainda mais para que o valor estimado seja o mais próximo do real (TUBINO, 2009).

Com o constante crescimento do setor de bebidas no Brasil e com as grandes mudanças que ocorreram por causa da pandemia de COVID-19, o estudo de previsões por métodos de séries temporais pode auxiliar no planejamento de ações mais efetivas, e tomadas de decisão mais eficazes, de acordo com as necessidades observadas nas análises realizadas.

O presente trabalho busca comparar métodos de previsão a partir das séries temporais de acordo com cada situação aplicada, além de estimular o estudo para entender melhor os impactos da pandemia de COVID-19 sentidos por esse setor industrial.

2 REFERENCIAL TEÓRICO

2.1 SÉRIES TEMPORAIS

Segundo BROCKWELL & DAVIS (2002), qualquer conjunto de observações X_t, onde cada observação é registrada em um instante específico do tempo *t*, é chamado de série temporal. Uma série temporal pode ser classificada como discreta ou contínua: as discretas são conjuntos de tempos discretos em que as observações feitas (ex. temperatura diária de uma determinada cidade); as contínuas são obtidas de forma contínua a partir de observações feitas em um intervalo de tempo de tempo $T = [t_1, t_2]$ onde t_1 e t_2 são instantes de tempo quaisquer (BROCKWELL; DAVIS, 2002).

A partir de uma série temporal, é possível definir objetivos de acordo com a necessidade desejada. MORETTIN & TOLOI (2006) e EHLERS (2007) explanam os principais objetivos para buscar-se o melhor modelo para uma série:

- fazer previsões de valores futuros a curto ou longo prazo;
- descrever o comportamento da série, identificando as suas principais componentes e outliers, e construindo gráficos para análise;
- entre outras informações que se pode obter.

Uma série temporal pode ser classificada como estacionária ou não-estacionária; também podendo conter uma estrutura interna que pode ser dividida, sendo composta por componentes de sazonalidade, tendência e resíduos (EHLERS, 2007).

2.1.1 Autocovariância e autocorrelação

A autocovariância é a covariância de uma variável com ela mesma, sendo vista como a covariância de uma variável com a sua defasagem. Dessa forma, a autocovariância é definida como:

$$cov_j(t) = cov(y_t, y_{t-j}) \tag{2.1}$$

Onde y_t é a variável aleatória e j é a sua defasagem. E sua função é definida como:

$$\gamma_{jt} = E[(y_t - \mu_t)(y_{t-j} - \mu_{t-j})] \tag{2.2}$$

Onde γ_{jt} é a função de autocovariância.

Já a autocorrelação é uma medida que mostra o quanto uma variável afeta as suas defasagens. Uma autocorrelação perfeita tem valor 1 e o seu oposto tem valor -1.

2.1.2 Estacionariedade

A estacionariedade está relacionada com a estabilidade de um processo estocástico, podendo, então, as séries temporais serem estacionárias ou não-estacionárias. Segundo MORETTIN & TOLOI (2006), geralmente supõe-se que uma série é estacionária, ou seja, que ao longo do tempo ela é desenvolvida aleatoriamente em torno de uma média constante, mostrando assim um equilíbrio estável. Mas, na realidade, isso não é verdade, pois a maior parte das séries encontradas mostram de alguma forma uma não-estacionariedade.

Pelo teorema de Wold, entende-se que toda série estacionária pode ser decomposta em um modelo simples de séries temporais (FISCHER, 1982). Assim, baseado no teorema, uma série não estacionária pode não ser verdadeiramente o processo gerador dos dados.

De acordo com BUENO (2011), no processo estocástico, é possível verificar a estacionariedade a partir de três propriedades fundamentais:

1. A média do processo não varia ao longo do tempo;

$$E(Y_t) = \mu \tag{2.3}$$

Onde μ é uma média constante.

2. A variância é constante ao longo do tempo;

$$Var(Y_t) = E[Y_t - E(Y_t)]^2 = \gamma_0 \tag{2.4}$$

Onde a variância é constante e finita.

3. A covariância entre duas observações não depende do tempo, mas apenas da distância entre essas observações.

$$cov(Y_t, Y_{t-j}) = E[Y_t - E(Y_t)][Y_{t-j} - E(Y_{t-j})] = \gamma_j \tag{2.5}$$

Onde $j = 1, 2, 3, \ldots, T-1$.

2.1.3 Sazonalidade

A sazonalidade de uma série é representada por um comportamento (de pico ou vale) que costuma se repetir em determinados períodos. Esse tipo de padrão, geralmente, é observado em séries anuais longas, e podem ser bruscos ou ter representações mais suaves.

A componente sazonal pode ser pode ser aditiva, apresentando flutuações sazonais constantes e independentes do nível global da série, ou pode ser multiplicativa, onde suas flutuações sazonais são dependentes do nível global da série (EHLERS, 2007).

2.1.4 Tendência

Representa um movimento geral no decorrer do tempo, podendo ser um crescimento dos dados, uma estabilidade ou um decrescimento. Uma tendência apresentada em uma série

temporal é facilmente identificada visualmente a partir do seu gráfico, sendo uma modificação a longo prazo no nível da série (PAL; PRAKASH, 2017).

De forma simples, podemos descrever a tendência linear de uma série como:

$$x_t = \alpha + \beta t + \varepsilon_t \tag{2.6}$$

Onde α e β são as constantes a serem estimadas no tempo t e ε_t é um ruído branco. A mudança de nível da série é dada por $\beta = m_t - m_{t-1}$, sendo m_t o nível médio da série (EHLERS, 2007).

2.1.5 Ruído Branco

O ruído/resíduo surge depois que o modelo de uma série temporal estacionária é ajustado. Segundo EHLERS (2007), os resíduos são a parte não explicada da série. Então, podemos descrever o resíduo pela equação:

$$\varepsilon_t = X_t - X'_t \tag{2.7}$$

Onde X_t é a série estacionária, X'_t é o modelo ajustado e ε_t é o resíduo da série.

Uma série, de forma ideal, será considerada um ruído branco se suas variáveis $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ são independentes e identicamente distribuídas $(i.i.d.)$, ou seja, todas as variáveis possuem média zero, variância constante e não existe autocorrelação para qualquer *lag* – quantidade fixa de tempo passada – (BUENO, 2011). Sendo assim, um ruído branco é uma série bem-comportada.

2.2 MODELOS CLÁSSICOS DE PREVISÃO

2.2.1 Modelos de Suavização Exponencial

Nesta subseção, serão apresentados dois tipos de modelos clássicos de séries temporais baseados em de suavização exponencial: o modelo simples e o modelo de Holt-Winters.

2.2.1.1 Modelo de Suavização Exponencial Simples (SES)

É utilizado em séries que não apresentam nem tendência e nem sazonalidade. O modelo básico considera $x_t = \mu_t + \varepsilon_t$, onde μ_t é a média não-estacionária e ε_t é o ruído branco. O nível da série a_t é uma estimativa da média μ_t, sendo definido como a média ponderada do valor atual x_t e do valor anterior do nível a_{t-1}. Então, temos:

$$a_t = \alpha x_t + (1-\alpha) a_{t-1}\ , \textit{onde}\ 0 < \alpha < 1 \tag{2.8}$$

Se α tende a 1, então temos pouca suavização (a previsão depende das observações mais recentes) e, consequentemente, o nível estará próximo da série x_t. Neste caso, o peso do ruído para a série temporal será pequeno, pois a variância do ruído σ_ε^2 é bem menor do que as

mudanças que ocorrem na média da série. Já se α tende a 0, haverá muita suavização (a previsão depende de muitas observações passadas) e a variância do ruído σ_ε^2 é bem maior do que as mudanças na média. Logo, o valor de α deve ser escolhido de forma que reflita a influência das observações passadas na previsão, fazendo com que o erro de uma previsão um passos à frente seja minimizado (BUENO, 2011; EHLERS, 2007)

As previsões são médias ponderadas das observações passadas com pesos que decaem exponencialmente a medida em que as observações ficam mais distantes, ou seja, quanto mais recente uma observação, maior é o peso associado (BUENO, 2011).

Então, a série suavizada, em temos de previsão, é dada por:

$$y_{t+1} = y_t + \alpha(x_{t+1} - y_t) \tag{2.9}$$

Onde x_{t+1} é a observação no período $t+1$ e y_{t+1} é a previsão um período a frente.

O modelo SES é de fácil entendimento e tem grande flexibilidade na variação da constante de suavização α, e por esses e outros motivos ele é tão utilizado, apesar da grande dificuldade para determinar o melhor valor para α (MORETTIN; TOLOI, 2006).

2.2.1.2 Modelo de Suavização Exponencial de Holt (SEH)

Modelo utilizado para analisar séries temporais que apresentam nível, tendência e resíduo aleatório. Além da constante de suavização α que afeta o nível da série, no modelo de SEH, é introduzida a constante β que afeta a tendência da série, mostrada na equação (*2.10)* (MORETTIN; TOLOI, 2006; VERÍSSIMO et al., 2013).

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{2.10}$$

Onde b_t é a tendência da série e β é a constante de suavização da tendência.

A previsão do método SEH é representada por

$$y_{t+k} = a_t + kb_t\,, \forall k > 0 \tag{2.11}$$

Onde y_{t+k} é a previsão e k é o número de passos à frente que se quer prever (MORETTIN; TOLOI, 2006).

2.2.1.3 Modelo de Holt-Winters (HW)

Similar ao método modelo simples, o método HW representa uma expansão SES. considerando não apenas o nível da série, mas possivelmente também considerando uma tendência e sazonalidade. Dependendo das características existentes na série temporal, as equações da série, cada uma associada a uma constante de suavização, podem ser diferentes, pois o método divide as séries temporais sazonais nas categorias: sazonalidade com uma

componente multiplicativa e sazonalidade com uma componente aditiva (MORETTIN; TOLOI, 2006).

- Multiplicativa: quando as amplitudes tendem a ter um crescimento ao longo do tempo (EHLERS, 2007). Assim, as equações nesse caso são:

$$a_t = \alpha\left(\frac{x_t}{s_{t-p}}\right) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{2.12}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \tag{2.13}$$

$$s_t = \gamma\left(\frac{x_t}{a_t}\right) + (1-\gamma)s_{t-p} \tag{2.14}$$

Onde a_t, b_t e s_t são o nível, inclinação e efeito sazonal, respectivamente; α, β e γ são os coeficientes de suavização de cada equação; p é o passo anterior do efeito sazonal; e x_t é o valor corrente.

- Aditiva: já para o modelo com sazonalidade aditiva, as equações de nível e de sazonalidade sofrem algumas modificações:

$$a_t = \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \tag{2.15}$$

$$s_t = \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \tag{2.16}$$

Para a previsão, também existem duas formas que calculá-la dependendo do efeito sazonal. Para um efeito multiplicativo, a previsão é dada por:

$$X_{t+k} = (a_t - kb_t)s_{t+k-p}\,, k = 1, 2, \ldots \tag{2.17}$$

Onde X_{t+k} é a previsão e k é o horizonte de tempo.

Já para um efeito aditivo da sazonalidade, a previsão é dada por:

$$X_{t+k} = a_t + kb_t + s_{t+k-p} \tag{2.18}$$

Como no modelo de suavização exponencial simples, os parâmetros α, β e γ são fundamentais para realizar a previsão. Devem ser escolhidos de forma que minimize a soma dos quadrados dos erros da previsão depois da aplicação do modelo (BROCKWELL; DAVIS, 2002).

2.2.2 Modelos ARIMA

Os modelos ARIMA podem ser usados tanto em séries estacionárias como em não-estacionárias. Assim, passos diferentes serão tomados a depender de como a série se encontra. Para séries estacionárias, pode-se buscar o modelo ARMA que melhor se adeque aos dados. Caso a série seja não-estacionária, primeiro devemos transformá-la em estacionária para, então,

encontrar o seu melhor modelo, de maneira que podemos considerar a classe de processos ARIMA (BROCKWELL; DAVIS, 2013).

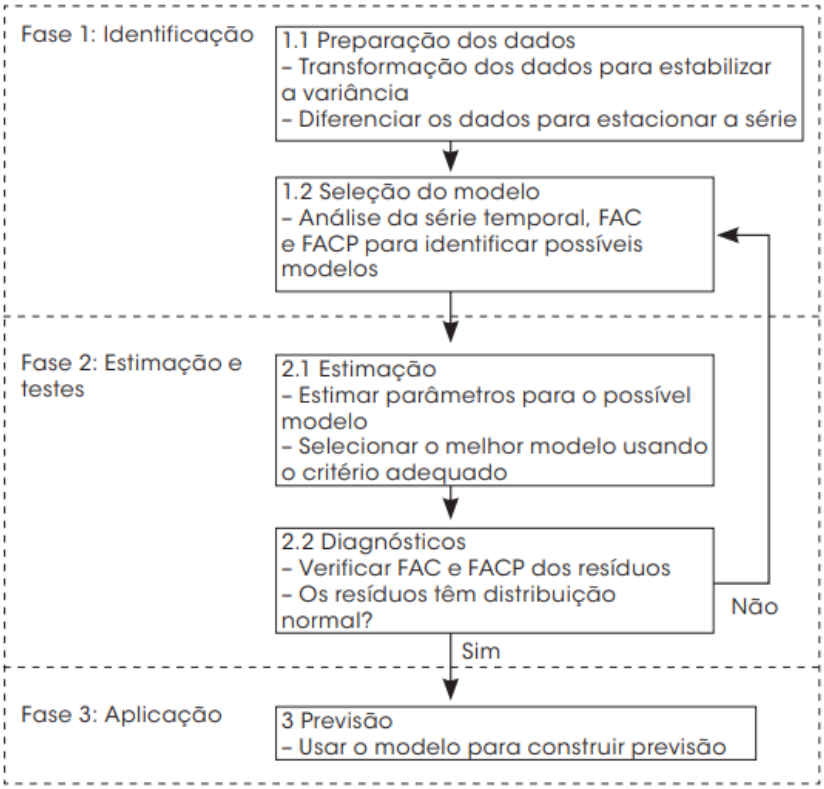
Segundo BOX & JENKINS (1976) e MORETTIN & TOLOI (2006), a construção do modelo é baseada em um ciclo interativo de passos que contém alguns estágios: identificação, estimação e diagnóstico, e previsão, como mostra a .

O estágio de identificação traz as etapas de diferenciação da série, se necessário, a análise das funções de autocorrelação – dada por um gráfico que mostra a correlação de uma variável contra a sua defasagem, mostrando se há ou não alguma relação entre a observação presente e a passada (BUENO, 2011) – e autocorrelação parcial – gráfico que mostra a correlação entre duas variáveis puras, eliminando qualquer correlação implícita que exista entre as duas observações – e demais critérios para que seja selecionado o modelo.

O estágio de estimação é onde os parâmetros do melhor modelo identificado são estimados e a análise dos resíduos é realizada. Ao final desse estágio, os resíduos devem se assemelhar a um ruído branco. Os dois primeiros estágios podem ser realizados inúmeras vezes até que o modelo adequado seja encontrado e satisfaça as condições e, de fato, represente o processo que gera a série (MORETTIN; TOLOI, 2006; ZHANG, 2003).

O último estágio é o de previsão, que se mostra como o objetivo principal do estudo e só é realizado quando as etapas anteriores são satisfatórias. Por fim, um bom modelo é aquele em que os seus resíduos se assemelham ao ruído branco, ou seja, que apresentam média zero, variância constante e variáveis sem autocorrelação.

Figura 1 – Etapas da abordagem Box-Jenkins



Fonte: Formigoni Carvalho Walter et al. (2013)

2.2.2.1 Modelo ARMA

Primeiro, para a aplicação dos modelos AR, MA ou ARMA, a série temporal considerada deve ser estacionária ou estar em sua forma estacionária. Nos modelos autorregressivos de ordem p – $AR(p)$ – a variável de interesse será prevista usando uma combinação linear dos valores anteriores da mesma, representando uma regressão da variável contra ela mesma. Ou seja, num AR de 1ª ordem é considerado o efeito, ou autocorrelação, que existe no período imediatamente anterior ao atual. Matematicamente, pode-se observar a equação (2.19).

$$\hat{X}_t = \sum_{i=1}^{p} \phi_i \cdot X_{t-i} + \varepsilon_t \qquad (2.19)$$

Onde p é a ordem de autorregressão e ϕ é o coeficiente (peso) (MORETTIN; TOLOI, 2006; PAL; PRAKASH, 2017). Definindo agora o operador autorregressivo (ou operador de defasagem) de ordem p, temos:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (2.20)$$

Podendo ser escrito como: $\phi(B)\widehat{X_t} = \varepsilon_t$, onde $\hat{X}_t = X_t - \mu$ (BOX et al., 2015; MORETTIN; TOLOI, 2006).

Nos modelos de médias móveis de ordem q – $MA(q)$ – são acessadas as tendências erráticas ao longo do tempo, de forma que cria uma média que é constantemente atualizada, ou seja, o processo de médias móveis sempre estará associado aos erros do modelo. Além disso, de acordo com a importância atribuída as observações passadas, os pesos associados poderão ser diferentes (BUENO, 2011). Matematicamente, temos a equação:

$$Y_t = \mu + \sum_{i=0}^{q} \theta_i \cdot \varepsilon_{t-i} \tag{2.21}$$

Onde q é a ordem de defasagens, θ é o coeficiente e ε_{t-i} é o erro (BUENO, 2011).

De acordo com a série estudada, podemos utilizar um modelo autorregressivo, de médias móveis ou, muitas vezes, um modelo misto baseado nos dois parâmetros anteriores chamado $ARMA\ (p, q)$. Na prática, em grande parte das séries, um modelo misto nos dá a possibilidade de não ter um grande número de parâmetros, tornando assim a solução mais conveniente (MORETTIN; TOLOI, 2006). A partir das equações dos modelos autorregressivo e de médias móveis, o modelo $ARMA\ (p, q)$ é definido pela equação (2.22).

$$\phi(B)\hat{X} = \theta(B)\varepsilon_t \tag{2.22}$$

2.2.2.2 Modelo ARIMA

Diferente dos modelos da Seção 2.2.2.1, usados para descrever séries estacionárias, o modelo ARIMA é utilizado para descrever séries não-estacionárias, sendo uma generalização do modelo ARMA. Ou seja, o ARIMA de ordem (p, d, q) é apropriado para séries que contém efeitos de tendência, sazonalidade e variância (MORETTIN; TOLOI, 2006).

Então, um $ARIMA\ (p, d, q)$ – em que p é o parâmetro autorregressivo, d é a diferenciação e q é o parâmetro de médias móveis – pode ser definido a partir da equação (2.23) que leva em consideração a diferença do processo.

$$W_t = \alpha_1 W_{t-1} + \cdots \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{2.23}$$

Onde W_t é a série diferenciada (EHLERS, 2007).

Dessa forma, a equação equivalente é dada por:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \tag{2.24}$$

Nesse processo autorregressivo integrado de médias móveis, observa-se que X_t é não estacionário, visto que $\phi(B)(1 - B)^d$ tem d raízes no círculo unitário, sendo necessário d diferenças para tornar a séries estacionária (EHLERS, 2007).

Após o processo de diferenciação da série para transformá-la em uma série estacionária, e consequentemente, sem os efeitos de sazonalidade, tendência e nível, é possível aplicar os modelos nos resíduos.

2.2.2.3 SARIMA

Assim como o modelo ARIMA, o modelo SARIMA surge para descrever modelos de séries não-estacionárias, mas agora apresentando uma autocorrelação sazonal (BOX; JENKINS, 1976). O modelo SARIMA foi definido por BOX & JENKINS (1976) como uma generalização do modelo ARIMA, chamando-o de modelo ARIMA sazonal multiplicativo, conhecido como $SARIMA\ (p,d,q)(P,D,Q)$. Matematicamente, pode ser representado pela equação (2.25).

$$\phi(B)\varphi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\vartheta(B^s)\varepsilon_t \tag{2.25}$$

Onde $\Delta_s^D X_t = (1-B^s)^D X_t$ é a ordem de diferenciação sazonal, $\varphi(B^s)$ é o coeficiente sazonal autorregressivo e $\vartheta(B^s)$ é o coeficiente sazonal de médias móveis (BOX; JENKINS, 1976; MARTINEZ; SILVA, 2011).

2.3 APRENDIZADO DE MÁQUINA (ML)

Classicamente, diz-se que um programa de computador aprende com uma experiência E em relação a uma classe de tarefas T e medida de desempenho P, se o desempenho das tarefas em T, medido por P, é melhorado com E (MITCHELL, 1997). O aprendizado de máquina (i.e., *machine learning*) é um dos campos da Inteligência Artificial (AI) que tem como objetivo o desenvolvimento de técnicas de aprendizagem assistida por computador, assim como o desenvolvimento de sistemas que são capazes de aprender por conta própria, buscando automatizar a detecção de padrões em dados (MONARD; BARANAUSKAS, 2003; SHALEV-SHWARTZ; BEN-DAVID, 2014).

A aprendizagem de máquina pode se dividir em três tipos: aprendizagem supervisionada, aprendizagem não-supervisionada e aprendizagem por reforço.

- Aprendizagem supervisionada: Envolve um conjunto de dados de entrada e de saída que são usados como treinamento para que o algoritmo "aprenda" a partir dos padrões encontrados nos dados. Tem como objetivo criar um modelo que rotule para qual classe um dado pertence. Para valores discretos, será feita a classificação dos dados. Já para valores contínuos, será feita uma regressão (MONARD; BARANAUSKAS, 2003).
- Aprendizagem não supervisionada: No conjunto de dados, apenas os dados de entrada estão disponíveis para o algoritmo. Então, a aprendizagem ocorre a partir da análise

desses dados, da busca pela relação implícita entre eles e da identificação de padrões para que possam ser agrupados, formando clusters (MAHESH, 2020).

- Aprendizado por reforço: O algoritmo aprende a partir da tentativa e erro para encontrar a melhor solução para o problema, envolvendo o conflito de entender o funcionamento do ambiente externo. Dessa forma, o sistema toma decisões a partir da experiência acumulada de problemas passados que tiveram uma solução positiva (JANIESCH; ZSCHECH; HEINRICH, 2021). Esse modelo, por exemplo, pode ser aplicado a jogos de computador (SILVER et al., 2018).

Para cada categoria de aprendizagem de máquina existem algoritmos que se encaixam melhor de acordo com os dados e com o que se quer deles. Entre os algoritmos, temos Redes Neurais Artificiais, *Support Vector Machines* (SVM), *Naive Bayes*, entre outros.

2.3.1 Support Vector Machine (SVM)

Dado um conjunto de funções de entrada e saída, usado como um conjunto para treinamento do algoritmo, pode-se prever dados futuros a partir dos exemplos aprendidos nesse treinamento (LIMA, 2014). Sendo o SVM um modelo de aprendizagem supervisionada, ele tem o objetivo de encontrar a função $f(x)$ que melhor responda aos dados, de maneira que se tenha o menor desvio dos dados reais (MAIOR et al., 2016).

Esse modelo pode ser dividido em duas categorias: classificação, conhecido como *Support Vector Classification* (SVC), e regressão, que é conhecido como *Support Vector Regression* (SVR). O SVR é o método utilizado neste trabalho.

Nos problemas de classificação, para os quais o SVM foi inicialmente desenvolvido, o objetivo é que o algoritmo receba novos exemplos de classes diferentes e seja capaz de classifica-los a partir do aprendizado vindo dos dados de treinamento (GUNN, 1998). O SVC pode ser aplicado em dados lineares – que podem ser separados por uma reta –, chamado de classificação linear, e para dados não lineares, chamado de classificação não linear.

Já em problemas de regressão, o SVR pode ser utilizado para funções não lineares e séries temporais. É um método usado para encontrar uma função $f(x)$ não linear que crie suas saídas constantes e minimize os erros da previsão realizada durante o processo de treinamento do algoritmo (MAIOR et al., 2016; SOUSA, 2018). Então, isso quer dizer que a ideia seria resolver um problema de otimização quadrática e convexa a partir da utilização das condições de Karush-Kuhn-Tucker (KKT), de maneira que ao final se tenha um ótimo global (MAIOR, 2017).

O objetivo do SVM é a busca por uma relação entre os dados de entrada e de saída e, dessa forma, encontrar uma solução ótima. A partir disso, a ideia é encontrar um hiperplano que represente da melhor forma o processo dos dados entre os dados de entrada e saída (MAIOR, 2015, 2017). A equação do hiperplano é definida pela equação (2.26)

$$f(x) = w^T x + b \tag{2.26}$$

Onde x são os dados de entrada e w^T e b são os coeficientes que serão determinados – a partir das equações (2.27) e (2.28) que mostram a função de risco regularizado.

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N}\psi_\varepsilon(d_i, f_i) + \frac{1}{2}\|w\|^2 \tag{2.27}$$

$$dado\ que\ \psi_\varepsilon(d_i, f_i) = \begin{cases} |d_i - df_i| - \varepsilon\,, & se\ |d_i - df_i| \geq \varepsilon \\ 0\,, & caso\ contrário \end{cases} \tag{2.28}$$

Onde d_i é o valor da variável – dado bruto – e f_i é o valor previsto para o mesmo período.

De acordo com FACELI et al. (2011), a realização de regressões não lineares pode ter a necessidade do uso das funções Kernel, tornando o algoritmo mais eficiente.

As funções Kernel dão a possibilidade de mapear os objetivos em espaços hiper dimensionais, onde a função linear mais regular e com baixo erro de treinamento pode ser encontrada (HOLFMANN, 2006). A escolha do Kernel para o algoritmo, em relação ao desempenho e a eficiência, depende diretamente do problema que está sendo modelado, mas pode se dividir em linear, polinomial, sigmoide e *radial basis function* (RBF).

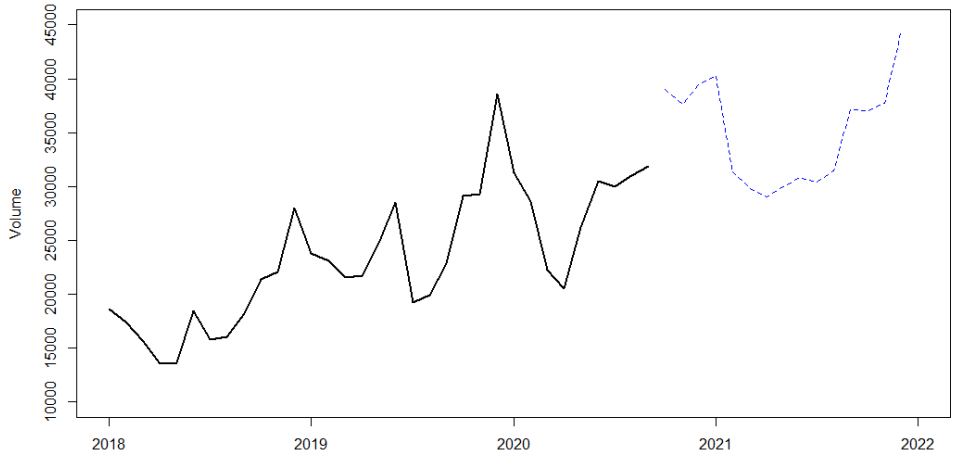
3 METODOLOGIA

3.1 APLICAÇÃO DOS MODELOS

Utilizando os dados da operação e considerando as especificidades de cada curva, como tendência e sazonalidade, foram usados dois modelos clássicos para análise de séries temporais: os modelos de HW e ARIMA. Adicionalmente, um modelo de *Machine Learning*, o SVM, foi utilizado.

Para todos os modelos, os dados foram divididos em duas partes: 70% destes foram utilizados como dados de treinamento e 30% como dados para teste do modelo. As performances dos modelos foram avaliadas comparando o valor real com a previsão considerando apenas os dados de teste. A primeira parte (dados de treinamento) consideram dados de janeiro de 2018 até setembro de 2020, enquanto a segunda parte (teste) considera dados de outubro de 2020 até dezembro de 2021. A Figura 2 apresenta o conjunto de dados onde a linha preta contínua representa os dados de treinamento enquanto a linha azul pontilhada os dados de teste.

Figura 2 – Série do volume total de bebidas dividida em treinamento e teste



Fonte: autor (2022)

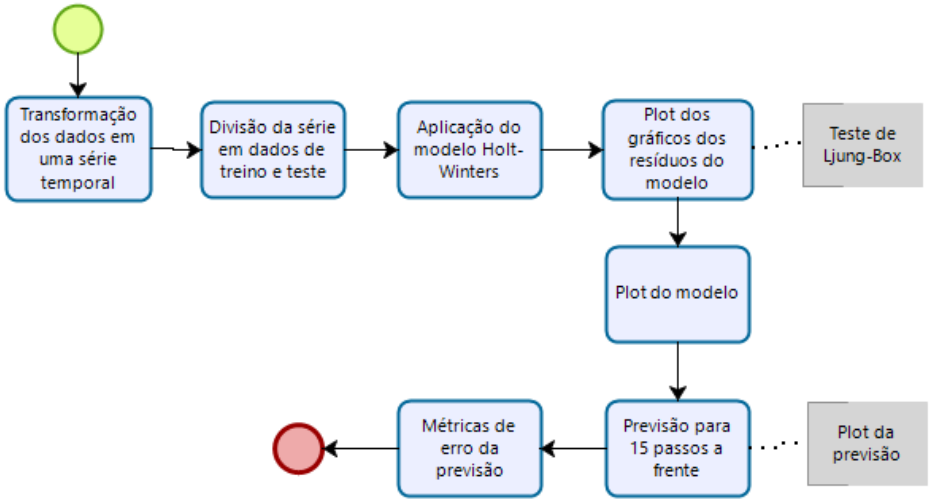
Além da série de volume total, os modelos também foram aplicados em séries estratificadas por produtos, além das séries dos canais de vendas. As previsões feitas são de um a 15 passos à frente (meses de teste), usadas como forma de avaliar o melhor modelo escolhido. A aplicação de cada modelo, em cada uma das séries utilizadas, foi feita de forma metódica, levando em consideração os fluxogramas do processo indicados nas Figura 3-5.

3.1.1 *Modelo de Holt-Winters*

O modelo de Holt-Winters foi aplicado a partir da função $HoltWinters()$ do pacote $forecast$ (HYNDMAN et al., 2020; HYNDMAN; KHANDAKAR, 2008) no RStudio, utilizando uma frequência de 12 meses, onde são obtidos os parâmetros de suavização α, β e γ que indicam, respectivamente, os coeficientes do nível da série, da tendência e da sazonalidade. Após a aplicação do modelo, foram feitos o teste $shapiro.test$, disponível como função *built-in* do R, que mostra aderência dos resíduos a normalidade e o teste *Ljung-Box*, realizado pela função $Box.test()$ do pacote $tseries$ (TRAPLETTI; HORNIK, 2019), para observar se há correlação dos dados. Além destes, também foram observados seus gráficos de ACF e PACF.

Uma vez selecionado o modelo com melhor ajuste, uma previsão até 15 passos à frente, de outubro de 2020 até dezembro de 2021, foi realizada e, por fim, foram analisadas as métricas de erro da previsão em comparação com a série real.

Figura 3 – Fluxograma de aplicação do modelo Holt-Winters



Fonte: autor (2022)

3.1.2 *Modelo ARIMA*

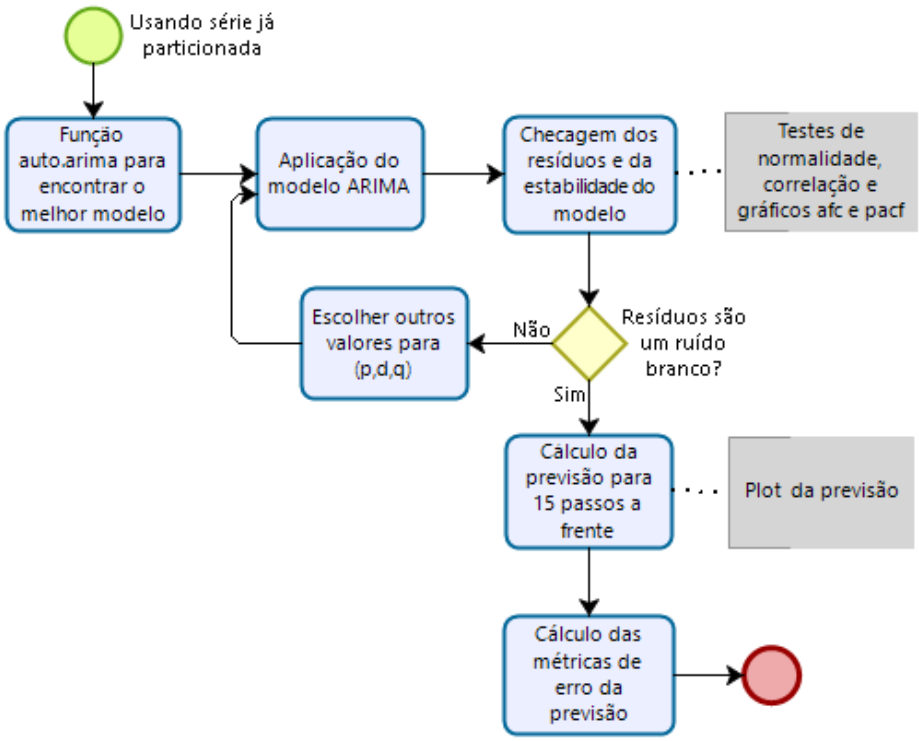
No modelo ARIMA, foi usada a função $auto.arima()$, do pacote $forecast$, para verificar qual a melhor ordem das variáveis do modelo levando em conta o menor coeficiente AIC. Dependendo do comportamento da série, podemos ter o modelo com as variáveis autorregressiva, de diferenciação e de médias móveis (p, d e q, respectivamente), mas também podemos ter, além das variáveis do modelo ARIMA, as variáveis de sazonalidade do modelo, indicadas por (P, D, Q) e o período sazonal encontrado.

Os gráficos ACF e PACF da série foram criados para conferir o resultado da função $auto.arima()$, além de também ser usada a função $ndiffs()$, do pacote $forecast$ para

confirmar o número de diferenciações necessárias para que a série se tornasse estacionária. Após o uso da função $auto.arima()$, para buscar os melhores parâmetros do modelo, a função $arima()$, do pacote $forecast$, foi utilizada respeitando todas as particularidades da série. Em seguida, verificou-se com o teste $shapiro.test$ se os resíduos seguem uma distribuição normal, se há correlação entre os dados com o teste $Ljung-Box$, se a série é estacionária e os gráficos ACF e PACF foram analisados.

Depois de confirmar que os resíduos são um ruído branco, foi calculada a previsão até 15 passos à frente e as métricas de erro entre a previsão e a série real foram analisadas.

Figura 4 – Fluxograma de aplicação do modelo ARIMA



Fonte: autor (2022)

3.1.3 Modelo Support Vector Machine (SVM)

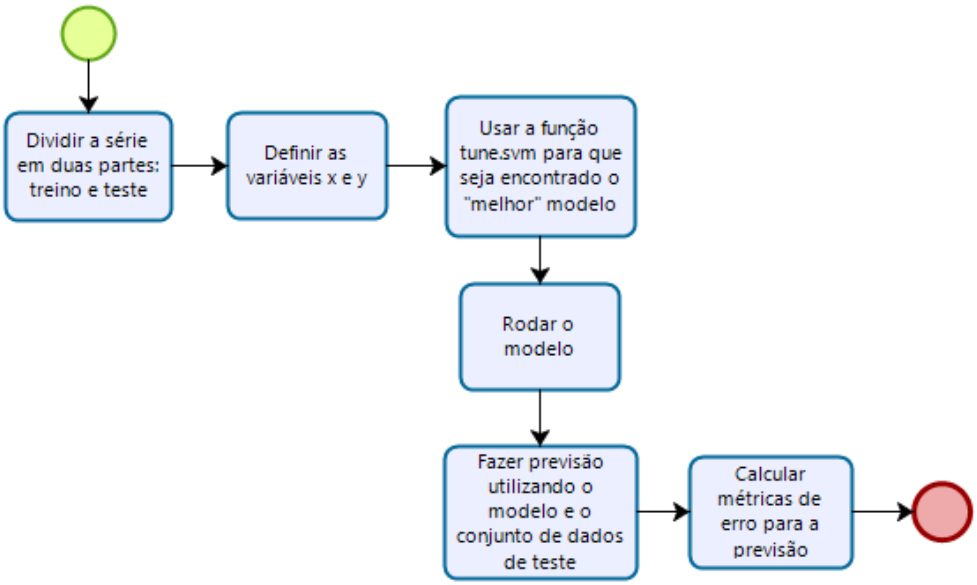
No modelo SVM, todas as séries foram ajustadas de maneira que o valor x da previsão y – para um período a frente – é sempre o valor da previsão y imediatamente anterior – y_{t-1}–, ou seja, x sempre será o lag_1 de y. Dessa forma, o modelo é representado pela função $y = f(x)$, onde y é a variável dependente de x.

Para que o melhor modelo fosse escolhido, a função $tune.svm()$ – que tem o objetivo de retornar os parâmetros otimizados para serem utilizados no modelo – do pacote $e1071$

(MEYER et al., 2021) do RStudio, foi utilizada. Após seu resultado, foi utilizada a função $svm()$, também do pacote $e1071$, para que o modelo fosse treinado.

Com o modelo treinado, utilizou-se o conjunto de dados de teste e o modelo escolhido para que a previsão fosse calculada 15 passos à frente. Por fim, as métricas de erro entre os valores previstos e os reais foram calculadas e analisadas.

Figura 5 – Fluxograma de aplicação do modelo SVM



Fonte: Autor (2022)

3.2 MÉTRICAS DE ERRO

Uma etapa importante da análise da previsão é a verificação a acuracidade do modelo através das métricas de erro dessa previsão (BUENO, 2011). A avaliação dessas medidas de desempenho é dada a partir comparação feita entre o valor previsto e o real. Para que se possa calcular, é necessário que as observações sejam divididas em uma série para o treinamento do modelo e uma série para teste – série usada para previsão e verificação (BUENO, 2011).

Dessa forma, duas importantes medidas de desempenho para o modelo foram utilizadas no presente trabalho: RMSE – *root mean squared error* – e MAPE – *mean absolute percentual error*.

- RMSE: Métrica que representa a raiz quadrada do desvio padrão da diferença entre o valor real e o previsto da amostra. Primeiro é calculado o somatório dos quadrados das diferenças, logo após é calculada a média e, por fim, é calculada a raiz quadrada.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \quad (3.1)$$

Onde n é o tamanho da amostra, x_i é o valor real e $\hat{x}_i$ é o valor previsto.

- MAPE: Nos informa a taxa de erro em termos percentuais. Primeiro calcula-se o erro percentual absoluto, logo após calcula-se o somatório dos erros e divide pelo número total de elementos.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right| \cdot 100 \quad (3.2)$$

Para as duas medidas de desempenho, quando menor for o seu resultado, melhores são as previsões realizadas. Neste trabalho, para os cálculos das métricas de erro nos modelos de Holt-Winters e no ARIMA, a função $accuracy()$ do pacote $forecast$ foi utilizada; já para o modelo SVM foram utilizadas as funções $MSE()$ e $MAPE()$ do pacote $MLmetrics$ (YAN, 2016).

4 ESTUDO DE CASO

A empresa, os dados e os processos realizados na série de dados para a previsão de demanda do volume de bebidas são apresentados neste capítulo.

4.1 DESCRIÇÃO DA EMPRESA

A operação utilizada neste trabalho é um centro de distribuição direta (CDD) que faz parte de uma empresa do setor de bebidas e está localizada na cidade de Caruaru, agreste pernambucano. O portifólio de produtos comercializados pode ser dividido em três: (i) cerveja, que contém todas as cervejas oferecidas pela empresa; (ii) *non alcoholic beverage* (NAB), contendo todos os produtos não alcoólicos (ex. refrigerantes, sucos, água); e (iii) Marketplace, que contém todos os diversos produtos de empresas terceiras. Para este estudo, serão considerados apenas os produtos do tipo cerveja e NAB, que são efetivamente produzidos pela empresa estudada.

O seguimento de vendas é dividido em seis canais, personalizando o atendimento prestado de acordo com a categoria de cada ponto de venda (PDV). São eles: AS Rota (ASR), Central de bebidas, FRIO, SUB, TRAD e VIP.

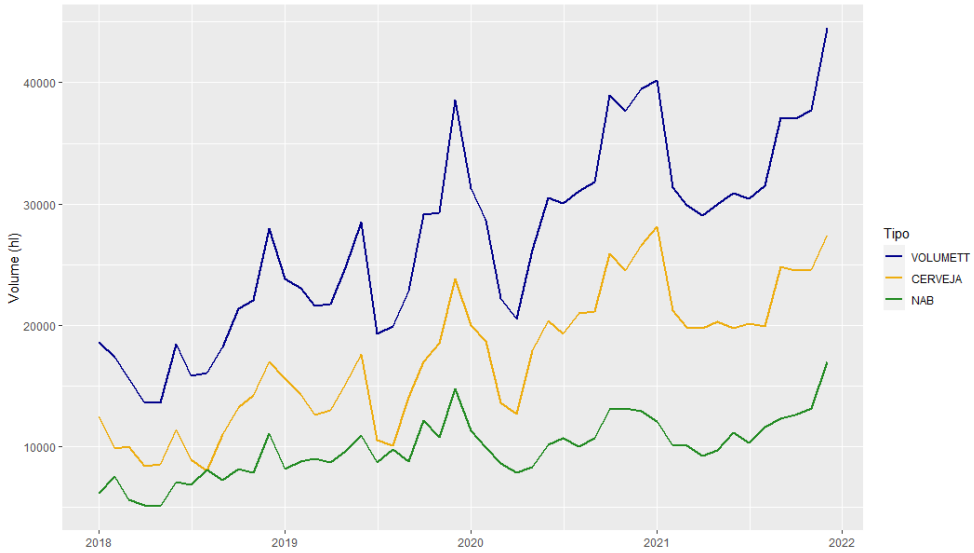
1. O AS Rota, autosserviço, atende os mercados que contém pelo menos um caixa, mas não atende as grandes redes de supermercado;
2. As centrais de bebidas atendem as cidades onde há dificuldade de o CDD manter um atendimento frequente, tornando-se franquias em parceria com estabelecimentos locais;
3. O FRIO atende todos os estabelecimentos que vendem a bebida gelada, para consumo no local. Por exemplo: bares e restaurantes;
4. O SUB atende todos os PDVs que vendem somente bebidas em grande quantidade, como as distribuidoras de bebidas;
5. O TRAD atende todos as mercearias/bodegas de bairro, onde geralmente não temos um *checkout* definido;
6. Por fim, o VIP engloba todos os estabelecimentos que têm um preço mais elevado, PDVs que influenciam a percepção de marca dos clientes. Por exemplo: pubs e restaurantes de shopping.

4.2 DESCRIÇÃO DOS DADOS

Os dados utilizados são de 48 meses, desde 2018 até 2021, do volume de bebidas vendido pela operação. O volume total foi dividido em dois tipos: Cerveja e NAB. Ainda, há

também uma subdivisão nos seis canais de vendas: ASR, Central de bebidas, FRIO, SUB, TRAD e VIP. As Figura 6 e Figura 7 mostram os gráficos dos dados dos tipos de produtos e dos canais, respectivamente. Assim, é possível obter uma melhor visualização do comportamento ao longo dos quatro anos observados.

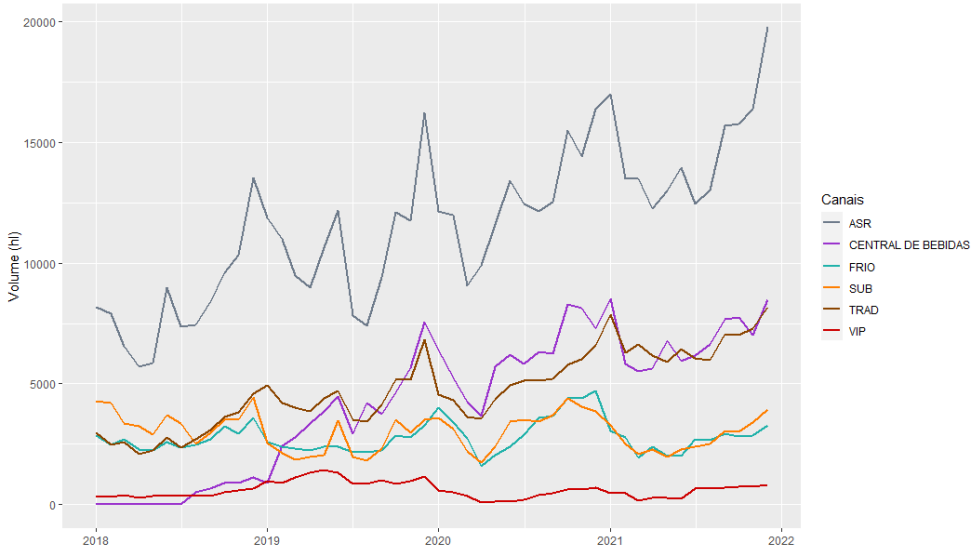
Figura 6 – Gráfico de volume total e volume por tipo de bebida



Onde VOLUMETT é o volume total. Fonte: autor (2022)

Apesar de a Figura 7 levar em consideração informações já apresentadas na Figura 6, é possível entender também as particularidades de cada canal, a sua curva no gráfico. Alguns canais mostram tendências de crescimento maiores do que outras, atingindo grandes picos e também grandes vales – como mostra a curva do canal ASR –, enquanto outras curvas mantém uma maior constância – como a curva do canal VIP.

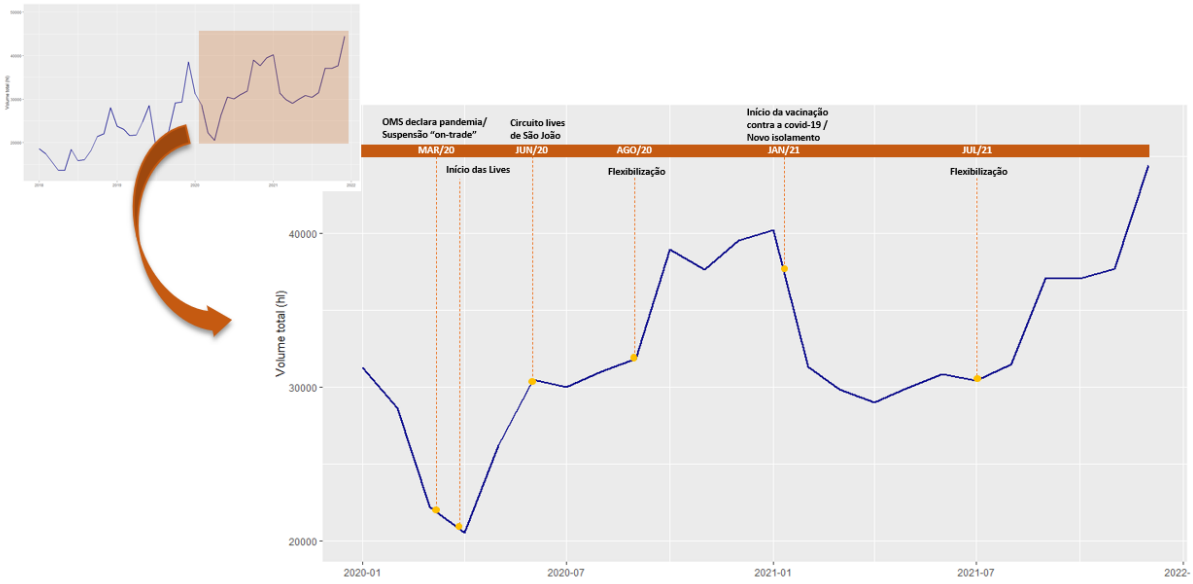
Figura 7 – Gráfico de volume por canal



Fonte: autor (2022)

É possível observar a tendência de crescimento ao longo dos anos tanto no de vendas de bebidas, mas com flutuações ao longo do tempo. Essas flutuações estão relacionadas a diversas mudanças no processo de vendas devido ao COVID-19. Desta forma, a Figura 8 apresenta uma linha do tempo relacionada a demanda requerida e eventos que possivelmente impactariam a mesma.

Figura 8 – Linha do tempo dos dados



Fonte: Autor (2022)

Percebe-se uma queda muito maior do que o normal entre os meses de março/2020 e abril/2020, que pode ser explicada pela declaração de pandemia de COVID-19 em 11/03/2020 pela OMS; e pela suspensão do funcionamento de estabelecimentos para consumo de alimentos e bebidas fora de casa, shows e festas no geral, de acordo com o decreto nº 48.809 do estado de Pernambuco que entrou em vigor a partir do dia 16/03/2020. Com o passar do tempo, a quarentena continuou e as pessoas tiveram que começar a viver os seus momentos de lazer em casa. Em 28/03/2020 houve a 1ª transmissão de um show ao vivo (*live*), das muitas que aconteceriam ao longo do ano de 2020. Como um show particular, da casa do artista para a casa do público, as *lives* foram crescendo e tendo o apoio, principalmente, de diversas marcas de cerveja.

Percebe-se na Figura 8 que as vendas de bebidas, no geral, começam a crescer rapidamente após um certo tempo de isolamento e isso pode estar diretamente ligado às festas particulares feitas em casa nos dias em que ocorriam as *lives*.

No início de 2021, pode-se perceber uma queda no volume maior do que o normal e isso pode ser explicado pelo novo isolamento causado pela segunda onda da COVID-19, além de fazer parte da sazonalidade dos dados que em dezembro tem um pico e logo após, um vale.

Sendo a região famosa pelas festas juninas e Caruaru conhecida como a capital do forró, principalmente por sua famosa festa de São João, podemos dizer que o maior São João da história, até o momento, em questão de volume de bebidas vendido, aconteceu em meio a uma pandemia.

A retomada do crescimento se mostra de forma mais lenta, pois apesar de o início da vacinação contra o novo coronavírus ter começado, em Pernambuco, no dia 18/01/2021, foi somente a partir do 2º semestre que as pessoas com mais de 18 anos puderam começar o processo de vacinação. A partir disso, começaram as flexibilizações e a retomada do funcionamento de estabelecimentos e festas.

5 DISCUSSÃO DOS RESULTADOS

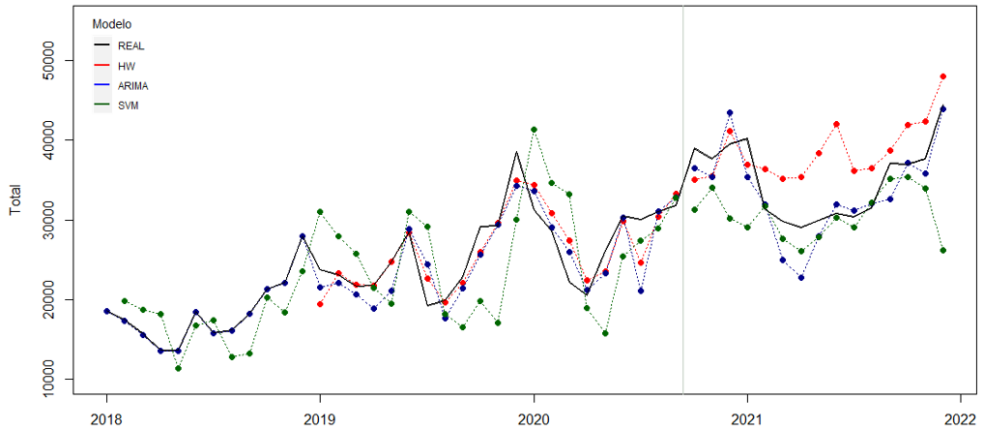
Nesta seção, a performance das três metodologias (i.e., HW, ARIMA e SVM) serão comparadas para previsão de demanda do volume total e de suas estratificações em linhas de bebidas e canais de vendas, como apresentado anteriormente. Os parâmetros de cada modelo encontram-se no apêndice A.

5.1 PREVISÕES

5.1.1 Série total de bebidas

A Figura 9 mostra o ajuste dos modelos à série de dados de treinamento, antes da linha vertical cinza, e as previsões realizadas para a série do volume total de bebidas do CDD, após a linha cinza. A linha contínua preta contém os dados reais e as linhas pontilhadas vermelha, azul e verde – HW, ARIMA e SVM, respectivamente – contém os ajustes e as previsões dos modelos.

Figura 9 – Ajustes e previsões do volume total por modelo



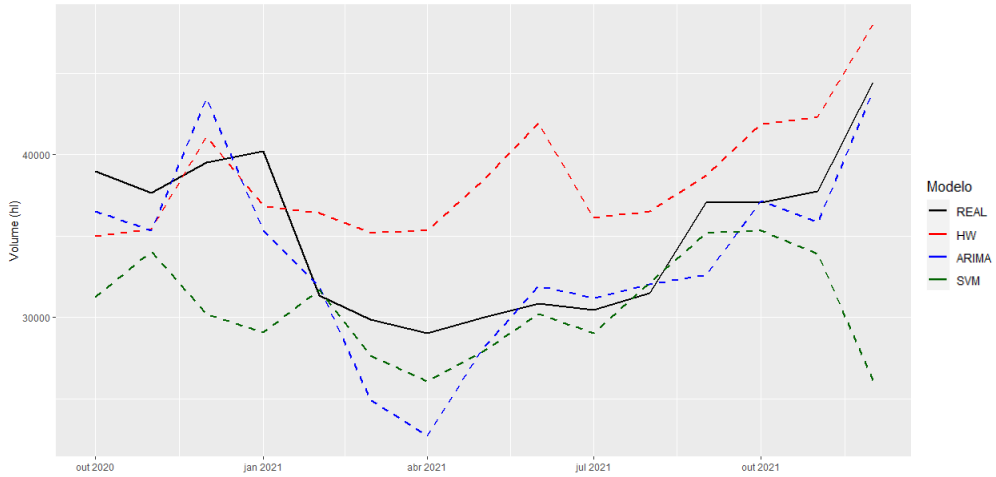
Fonte: autor (2022)

Em alguns casos, percebe-se um bom ajuste dos modelos no conjunto de treinamento, mas uma baixa performance no conjunto de teste, como por exemplo, no caso da curva do modelo HW. Apesar do exemplo da curva do HW parecer, em nenhum dos três casos é possível perceber um *overfitting* – quando o modelo se ajusta tão bem aos dados de treino que só consegue explica-los, errando consideravelmente as previsões de dados que não conhece ainda – presente fortemente.

Também não é possível caracterizar um *underfitting* – quando o modelo não consegue se ajustar bem nem aos dados de treino nem aos dados de teste – nas curvas dos modelos, visto que mesmo elas não tendo um ajuste perfeito para os dados de treino, ainda há um ajuste consideravelmente interessante. Em um bom modelo, busca-se um equilíbrio nos ajustes. O

modelo não precisa se ajustar excessivamente aos dados de treinamento, ele precisa se ajustar da melhor forma possível aos dados de teste, conseguindo, desta forma, explicar os dados que ele ainda não viu e não os dados que ele já conhece.

Figura 10 – Previsões do volume total por modelo



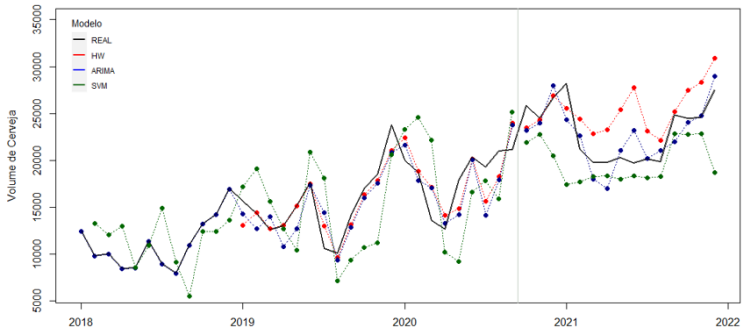
Fonte: autor (2022)

Baseado na Figura 10, visualmente, a curva de previsão que melhor acompanha o comportamento da curva de dados reais é a do ARIMA, trazendo alguns picos e vales semelhantes ao real. A curva do SVM também se aproxima, mas contém alguns *outliers* – como o último ponto da previsão – que se distancia bastante do valor real, podendo fazer com que o resultado de uma das medidas de desempenho utilizadas aumente consideravelmente, afetando a precisão do modelo. Por último, a curva de HW se mostra um pouco mais distante do real do que as outras curvas. Por ser um modelo de suavização, as quedas mais repentinas no volume podem não ser previstas pelo modelo, porque o algoritmo tenta suavizar e manter a regularidade do que acontecia antes desse pico ou vale repentino.

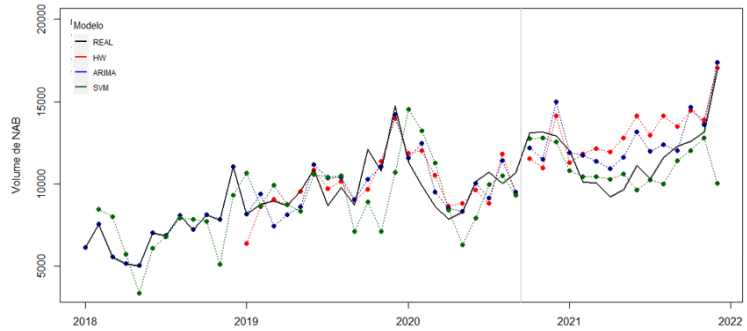
5.1.2 Estratificações por tipo de bebida e canal de vendas

Assim como na série de volume total, nas séries estratificadas é possível notar que o modelo ARIMA consegue se ajustar bem às curvas dos dados do conjunto de treinamento. Além dos ajustes serem notados visualmente, também é possível observar o desempenho do modelo durante o treinamento nas Tabela 1 e Tabela 2 que mostram as métricas de erro de cada curva, o que só comprova que o ARIMA é o modelo que tem um melhor ajuste. Entretanto, ter o melhor desempenho durante a fase de treinamento não garante um bom desempenho durante a fase de testes, sendo notável o desempenho mediano que o ARIMA teve na maioria das previsões para o conjunto de testes.

Figura 11 – Ajustes e previsões dos tipos de bebida por modelo



(a)



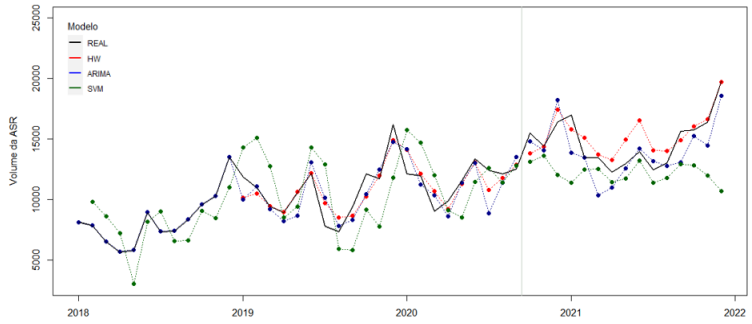
(b)

(a) Série do volume de cerveja (b) Série do volume de NAB. Fonte: autor (2022)

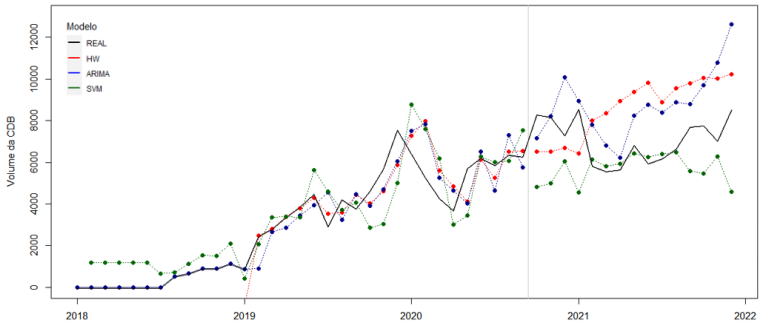
Diferentemente do ARIMA (i.e., melhor desempenho na fase de treinamento e um decaimento no conjunto de testes), o SVM não se mostrou com um ajuste tão interessante no treinamento do modelo, mas apresentou boas previsões para os dados de teste. Em grande parte das previsões das séries dos canais de vendas, o modelo teve o melhor desempenho, como por exemplo na Figura 12(f), série do volume do VIP, que teve um ótimo desempenho e ajuste aos dados de teste quase perfeito. Apesar dos grandes acertos do modelo, há também grandes erros em algumas séries, como nas séries do FRIO, Central e TRAD (Figura 12), com previsões em direção oposta a esperada.

Nas previsões de HW, é possível perceber que o modelo, assim como o SVM, não consegue explicar tão bem os dados de treinamento, mas também não consegue ter boas previsões para o conjunto de testes, podendo ser visto, muitas vezes, como um modelo *underfitting*. Em grande parte das séries, as previsões do modelo foram de baixa qualidade, visto que em algumas das séries tiveram erros enormes, como é possível notas na série do SUB ou do VIP da Figura 12. Mas mesmo com grandes erros, o modelo de HW conseguiu ter um desempenho satisfatório em algumas das séries, como na do TRAD, da Figura 12.

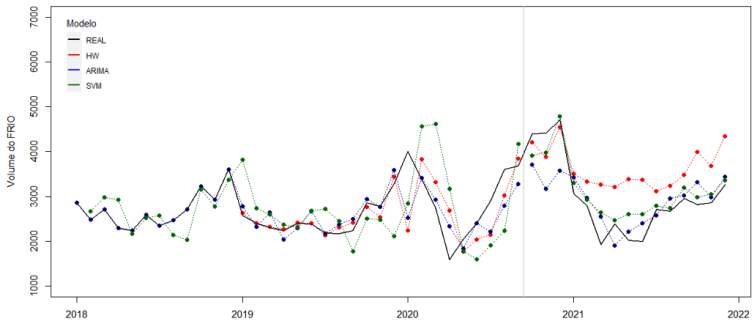
Figura 12 – Ajustes e previsões dos canais de venda por modelo



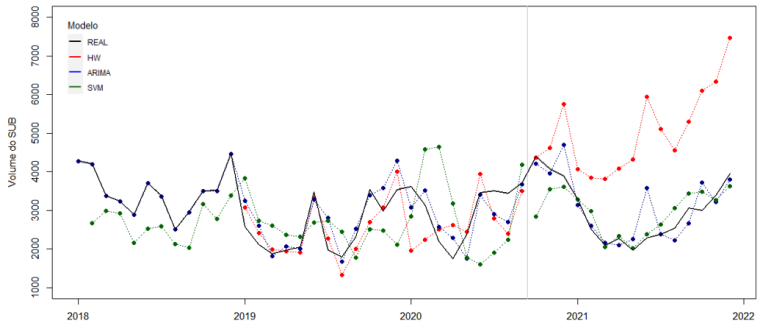
(a)



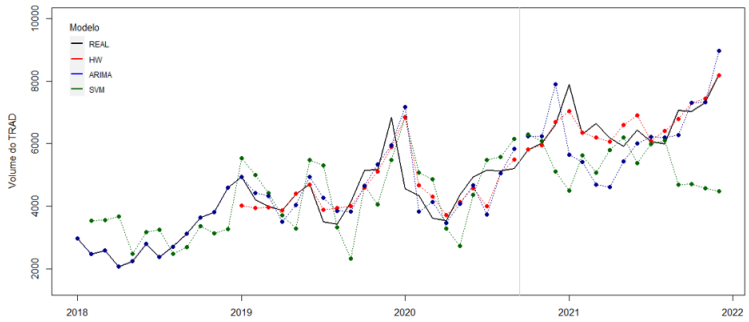
(b)



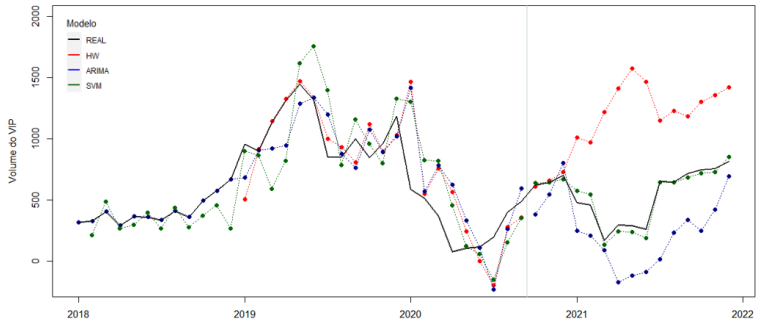
(c)



(d)



(e)



(f)

(a) Canal ASR (b) Central de Bebidas (c) FRIO (d) SUB (e) TRAD (f) VIP. Fonte: autor (2022)

5.2 MÉTRICAS DE ERRO

Dadas as previsões, a precisão dos modelos foi calculada a partir das métricas de erro RMSE e %MAPE. As Tabela 1 e Tabela 2 mostram os valores correspondentes ao conjunto de dados de treinamento; já as Tabela 3 e Tabela 4 mostram os valores que correspondem ao conjunto dados de teste. Elas apresentam os valores de acordo com a série, modelo e métrica utilizada, valendo ressaltar que os valores em negrito indicam os modelos com melhor desempenho e os valores sublinhados indicam os modelos com pior desempenho.

Tabela 1 – Métricas de erro dos ajustes das séries de volume total, de cerveja e de NAB para os dados de treinamento

	HW		**ARIMA**		**SVM**	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
TOTAL	2555,81	7,21	**2474,72**	**5,76**	5939,98	21,26
CERVEJA	1961,45	8,85	**1886,91**	**7,92**	4472,92	27,40
NAB	1179,42	9,12	**865,27**	**5,51**	1905,41	16,88
Média	1898,89	8,39	**1742,30**	**6,40**	4106,11	21,85

Fonte: autor (2022)

Aqui é possível observar de maneira quantitativa indícios visualizados nas figuras da Seção 5.1. De fato, o ARIMA é o modelo com melhor ajuste aos dados de treinamento, seguido do HW e SVM, tanto para a série total como para as estratificações.

Se os modelos fossem escolhidos baseados nas métricas de erros do conjunto de treino, certamente, pelas Tabela 1 e Tabela 2, o ARIMA seria escolhido como o melhor e o SVM como o pior. Mas como mencionado anteriormente, pouco adianta um excelente ajuste aos dados de treino caso a performance nos dados de teste seja ruim. De fato, de maneira esperada, observa-se considerável diferença que há entre os valores das tabelas com os desempenhos do teste e do treinamento, gerando reflexões sobre qual seria o modelo mais indicado.

Tabela 2 – Métricas de erro das previsões das séries dos canais de venda para os dados de treinamento

	HW		**ARIMA**		**SVM**	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
ASR	**1063,10**	7,55	1143,91	**6,89**	2503,24	22,09
CENTRAL	1047,33	23,34	**893,55**	**17,39**	1293,38	27,41
FRIO	541,53	12,36	**375,11**	**7,45**	744,70	21,55
SUB	619,35	17,24	**373,62**	**9,44**	1052,52	30,90
TRAD	686,54	9,75	**607,07**	**7,30**	1002,47	22,42
VIP	288,61	70,82	**243,76**	**49,08**	280,03	50,12
Média	707,74	23,51	**606,17**	**16,26**	1146,06	29,08

Fonte: autor (2022)

Na Tabela 3, se fosse levado em consideração apenas um dos valores de erro para avaliar o desempenho do modelo, é possível que a análise não fosse a mesma dependendo da métrica utilizada uma vez que não há consistência entre as duas métricas para alguns modelos, como por exemplo: (i) a série Total tem no modelo de HW o MAPE com o pior desempenho, mas o RMSE com um valor intermediário; (ii) analisando a mesma série, o modelo SVM mostra-se o contrário do HW, com o RMSE tendo o pior desempenho e o MAPE com um desempenho intermediário.

Essa não consistência pode se dar devido ao cálculo do RMSE, que por utilizar um somatório dos quadrados dos erros antes de calcular a média, faz com que os pesos que serão atribuídos a soma aumentem à medida em que os erros aumentam. Ou seja, havendo um *outlier* na série de dados, o peso dele será maior no cálculo.

Tabela 3 – Métricas de erro das previsões das séries de volume total, de cerveja e de NAB para os dados de teste

	HW		**ARIMA**		**SVM**	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
TOTAL	5400,30	14,65	**3121,24**	**8,85**	6639,83	11,77
CERVEJA	3495,24	23,42	**2022,04**	**7,22**	4383,22	13,67
NAB	2028,13	16,57	**1448,07**	11,37	1978,83	**8,67**
Média	3641,22	18,21	**2197,12**	**9,15**	4333,96	11,37

Fonte: autor (2022)

Para as séries total e de tipos de bebida, o modelo ARIMA obteve o melhor desempenho para as duas métricas de erro em duas das três séries, mas teria o melhor desempenho nas 3 séries se fosse levado em consideração somente o RMSE.

Para a Tabela 4, os resultados do RMSE e MAPE se mostraram mais consistentes do que anteriormente. O SVM se destaca como melhor modelo em três das seis séries de canais, e ainda melhora se olhássemos apenas para o MAPE, tendo o melhor desempenho em quatro das seis séries. Contrário ao SVM, o modelo de HW obteve o pior desempenho em quatro das seis séries. É possível notar também que nas séries em que o HW teve o melhor desempenho, o SVM teve o pior.

Tabela 4 – Métricas de erro das previsões das séries dos canais de venda para os dados de teste

	HW		**ARIMA**		**SVM**	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
ASR	**1263,35**	**7,28**	1591,71	8,00	3462,51	16,12
CENTRAL	2496,82	35,53	2183,69	27,09	**1288,62**	**12,57**
FRIO	854,22	29,96	558,15	14,18	**352,86**	**10,88**
SUB	2286,85	75,14	**471,24**	11,67	503,43	**10,52**
TRAD	**363,22**	**3,93**	1023,91	11,57	1831,44	19,28
VIP	719,97	169,71	348,16	65,98	**47,34**	**9,86**
Média	1330,74	53,59	**1029,48**	23,08	1247,70	**13,21**

Fonte: autor (2022)

O modelo de HW quando teve um bom desempenho nas previsões das séries de canais, os resultados obtidos foram bastante precisos, mas nas previsões em que obteve pior desempenho, ele foi bastante impreciso. O modelo ou tem grandes acertos ou tem grandes erros.

Para os resultados da Tabela 4, o modelo ARIMA foi o que se manteve mais estável em seus resultados, não estando com o pior desempenho em nenhuma das previsões, mas tendo o melhor desempenho apenas para o RMSE da série do canal SUB.

Finalmente, observando apenas as médias dos modelos, para as séries da Tabela 3 o ARIMA seria o modelo mais indicado para ser utilizado, mas para a Tabela 4 o melhor modelo fica entre o ARIMA e o SVM se a consistência das duas métricas de erro forem levadas em consideração. Se somente o RMSE for observado, o modelo ARIMA seria o mais indicado para a previsão das séries dos canais, mas se apenas o MAPE for levado em consideração, o modelo SVM seria o mais indicado.

6 CONCLUSÕES

Com as mudanças no mercado e nos hábitos e necessidades do consumidor, torna-se indispensável que as empresas utilizem metodologias para previsão da demanda, que além de estimar quantitativamente o que pode acontecer, traz reflexões e oportunidades de melhorias através da tomada de decisão.

O presente trabalho se propôs a, a partir de dados históricos de um CDD de bebidas, fazer uma análise através previsões e comparar modelos clássicos de séries temporais, modelo de HW e ARIMA, e um modelo de aprendizado de máquina, SVM. As performances dos modelos foram comparadas a partir de nove séries de dados, sendo uma série de volume total de bebidas, duas estratificadas em tipos de bebidas e seis em canais de venda.

As análises realizadas não trazem uma resposta definitiva, abrindo espaço para reflexões, visto que um único modelo não foi indicado como o melhor. Para cada série, obteve-se o que melhor se ajustou e teve bom desempenho. Apesar disso, dois modelos se destacaram, obtendo um grande número de acertos, o ARIMA e o SVM.

O ARIMA não esteve, em nenhuma das comparações, com o pior desempenho, ficando em sua maioria com um resultado intermediário. Já o SVM, esteve algumas vezes com o pior desempenho, mas também apresentou o melhor desempenho na maioria das séries de canais. O modelo de HW apresentou resultado satisfatório em duas das séries, mas teve grandes erros nas outras previsões realizadas.

Sendo assim, dependendo da medida de desempenho escolhida, pode-se entender qual o modelo mais indicado para a série analisada como, por exemplo: na série do volume do NAB se a medida RMSE for escolhida, o modelo mais indicado é o ARIMA, mas se o MAPE for escolhido, o SVM é a melhor alternativa.

Deste modo, para estudos futuros existem possibilidades de explorar esta pesquisa. Um ponto seria utilizar mais metodologias, além de considerar outros modelos de ML – como *Random Forest* e Redes Neurais Artificiais –, a fim de maiores comparações, podendo ter a oportunidade de haver um modelo que seja indicado para todos os casos. Uma outra sugestão para os próximos passos da pesquisa seria realizar uma estimação intervalar, possibilitando ainda mais análises sobre o modelo estudado, podendo ser feita através do uso do *bootstrap*, por exemplo.

REFERÊNCIAS

ABIA. **Associação Brasileira das Indústrias de Alimentação**. Disponível em: <https://www.abia.org.br/vsn/anexos/faturamento2019.pdf>. Acesso em: 20 mar. 2022.

BARZOLA-MONTESES, J. et al. T**ime series analysis for predicting hydroelectric power production: The ecuador case**. Sustainability (Switzerland), v. 11, n. 23, p. 1–19, 2019.

BOX, G. E. P. et al. **Time Series Analysis: Forecasting and Control**. 5th ed. ed. Hoboken: John Wiley & Sons, 2015.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis Forescasting and Control**. [s.l.] San Francisco, 1976.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. 2nd ed. ed. New York: Springer-Verlag New York, 2002.

BROCKWELL, P. J.; DAVIS, R. A. **Time Series: Theory and Methods**. New York: Springer Science & Business Media, 2013.

BUENO, R. DE L. DA S. **Econometria de Séries Temporais**. 2ª edição ed. São Paulo: Cengage Learning, 2011.

CONG, J. et al. **Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018**. International Journal of Environmental Research and Public Health, v. 16, n. 23, p. 4760, 27 nov. 2019.

CORTEZ, P.; DONATE, J. P. **Global and decomposition evolutionary support vector machine approaches for time series forecasting**. Neural Computing and Applications, v. 25, n. 5, p. 1053–1062, 22 out. 2014.

EHLERS, R. S. **Análise de Séries Temporais**. 4ª edição ed. Paraná: Laboratório de Estatística e Geoinformação Universidade Federal do Paraná, 2007.

FACELI, K. et al. **Inteligência artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora Ltda., 2011.

FANIBAND, Y. P.; SHAAHID, S. M. **Univariate Time Series Prediction of Wind speed with a case study of Yanbu, Saudi Arabia**. International Journal of Advanced Trends in Computer Science and Engineering, v. 10, n. 1, p. 257–264, 15 fev. 2021.

FISCHER, S. **Séries univariantes de tempo: metodologia de Box & Jenkins**. Porto Alegre, RS: Fundação de Economia e Estatística, 1982.

FORMIGONI CARVALHO WALTER, O. M. et al. **Aplicação de um modelo SARIMA na previsão de vendas de motocicletas**. Exacta, v. 11, n. 1, p. 77–88, 28 maio 2013.

GUNN, S. **Support vector machines for classification and regression**. University of

Southampton: Technical report, Image Speech & Intelligent Systems Group, 1998.

HOLFMANN, M. **Support Vector Machines - Kernel and the Kernel Trick**, 2006.

HYNDMAN, R. et al. **_forecast: Forecasting functions for time series and linear models_**R package version 8.13, , 2020. Disponível em: <https://pkg.robjhyndman.com/forecast/>

HYNDMAN, R.; KHANDAKAR, Y. **Automatic time series forecasting: The forecast package for R**_Jornal of Statistical Software_, , 2008. Disponível em: <https://www.jstatsoft.org/article/view/v027i03>

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. **Machine learning and deep learning.** Electronic Markets, v. 31, n. 3, p. 685–695, 8 set. 2021.

JUNIOR, O. C. et al. **O setor de bebidas no Brasil**. BNDES Setorial, 2014.

JUNIOR, O. C. **Panoramas Setoriais 2030 - Bebidas**. BNDES, 2017.

KRAJEWSKI, L. J.; RITZMAN, L. P.; MALHOTRA, M. K. **Administração de Produção e Operações**. 8ª ed. ed. [s.l.] Pearson Education do Brasil, 2009.

LEMOS, F. DE O. **Metodologia para seleção de métodos de previsão de demanda**. [s.l.] Universidade Federal do Rio Grande do Sul, 2006.

LIMA, J. G. B. **Uma aplicação de impacto social com aprendizagem de máquina**. [s.l.] Universidade Federal Rural de Pernambuco, 2014.

MAHESH, B. **Machine Learning Algorithms - A Review. I**nternational Journal of Science and Research (IJSR), v. 9, n. 1, 2020.

MAIOR, C. B. S. **Estimação de Residual Useful Life a partir de Empirical Mode Decomposition Support Vector Machine.** Universidade Federal de Pernambuco, 2015.

MAIOR, C. B. S. et al. **Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine.** IEEE Latin America Transactions, v. 14, n. 11, p. 4603–4610, 2016.

MAIOR, C. B. S. **Remaing Useful Life prediction via Empirical Mode Decomposition, Wavelets and Support Vector Machine**. [s.l.] Universidade Federal de Pernambuco, 2017.

MAKRIDAKIS, S. **Forecasting: its role and value for planning and strategy.** International Journal of Forecasting, v. 12, n. 4, p. 513–537, dez. 1996.

MARTINEZ, E. Z.; SILVA, E. A. S. DA. **Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model**. Cadernos de Saúde Pública, v. 27, n. 9, p. 1809–1818, set. 2011.

MEYER, D. et al. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien**R package version 1.7-6, , 2021. Disponível em: <https://cran.r-project.org/package=e1071>

MITCHELL, T. M. **Machine Learning**. [s.l.] McGraw-Hill Science/Engineering/Math, 1997.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre Aprendizado de Máquina. In: Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri, SP: Editora Manole Ltda, 2003. p. 89–114.

MONTGOMERY, D. C.; JOHNSON, L. A.; GARDINER, J. S. **Forecasting and Time Series Analysis**. New York: McGraw-Hill, 1990.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2ª edição ed. São Paulo: Editora Blucher, 2006.

NENNI, M. E.; GIUSTINIANO, L.; PIROLO, L. **Demand Forecasting in the Fashion Industry: A Review**. International Journal of Engineering Business Management, v. 5, p. 37, 1 jan. 2013.

PAL, A.; PRAKASH, P. K. S. **Practical Time Series Analysis: Master Time Series Data Processing, Visualization, and Modeling using Python**. [s.l.] Packt Publishing, 2017.

POLARY, J. H. B. **Indústria de Bebidas - Estudo Setorial**. FIEMA, 2021.

RADHIKA, Y.; SHASHI, M. **Atmospheric Temperature Prediction using Support Vector Machines**. International Journal of Computer Theory and Engineering, 2009.

ŠENKOVÁ, A. et al. **Time Series Modeling Analysis of the Development and Impact of the COVID-19 Pandemic on Spa Tourism in Slovakia**. Sustainability, v. 13, n. 20, p. 11476, 17 out. 2021.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning: From Theory to Algorithms**. [s.l.] Cambridge University Press, 2014.

SILVER, D. et al. **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play**. Science, v. 362, n. 6419, p. 1140–1144, 7 dez. 2018.

SOUSA, L. R. R. DE. **Utilização do aprendizado de máquina para o desenvolvimento de um modelo computacional para previsão de risco de dengue em Palmas - TO**. [s.l.] Centro Universitário Luterano de Palmas, 2018.

TRAPLETTI, A.; HORNIK, K. **tseries: Time Series Analysis and Computational Finance**R package version 0.10-47., , 2019.

TUBINO, D. F. **Planejamento e Controle da Produção: Teoria e Prática**. 2. ed. ed. São Paulo: Atlas, 2009.

VERÍSSIMO, A. J. et al. **MÉTODOS ESTATÍSTICOS DE SUAVIZAÇÃO EXPONENCIAL HOLT-WINTERS PARA PREVISÃO DE DEMANDA EM UMA EMPRESA DO SETOR METAL MECÂNICO**. Revista Gestão Industrial, v. 8, n. 4, 8 fev. 2013.

VIANA, F. L. E. **Indústria de bebidas alcoólicas**. Caderno Setorial ETENE, v. 117, 2020.

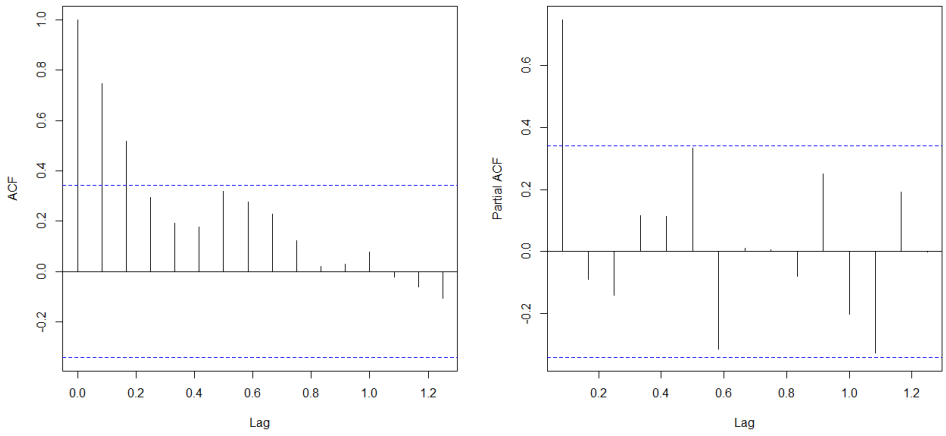
VIANA, F. L. E. **Indústria de bebidas não alcoólicas**. Caderno Setorial ETENE, v. 175, 2021.

YAN, Y. **MLmetrics: Machine Learning Evaluation Metrics**R package version 1.1.1, , 2016. Disponível em: <https://cran.r-project.org/package=MLmetrics>

ZHANG, G. P. **Time series forecasting using a hybrid ARIMA and neural model**. Neurocomputing, 2003.

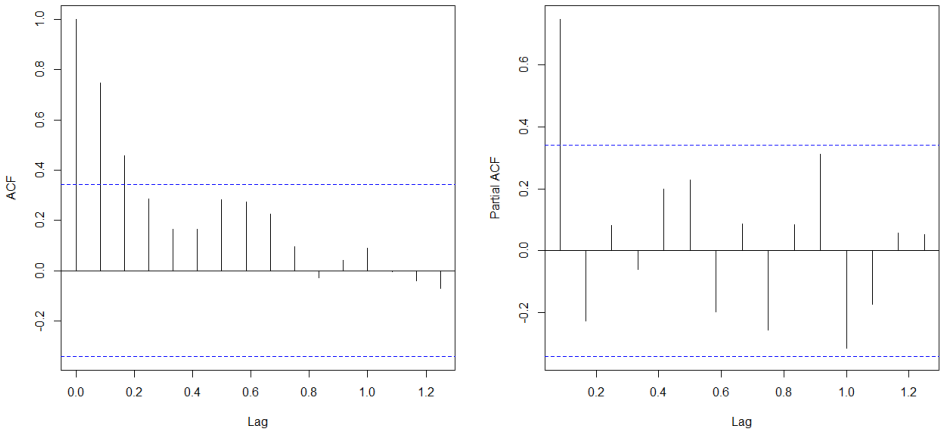
APÊNDICE A – Gráficos de ACF e PACF do modelo ARIMA

Figura 13 – Série do volume total de bebidas



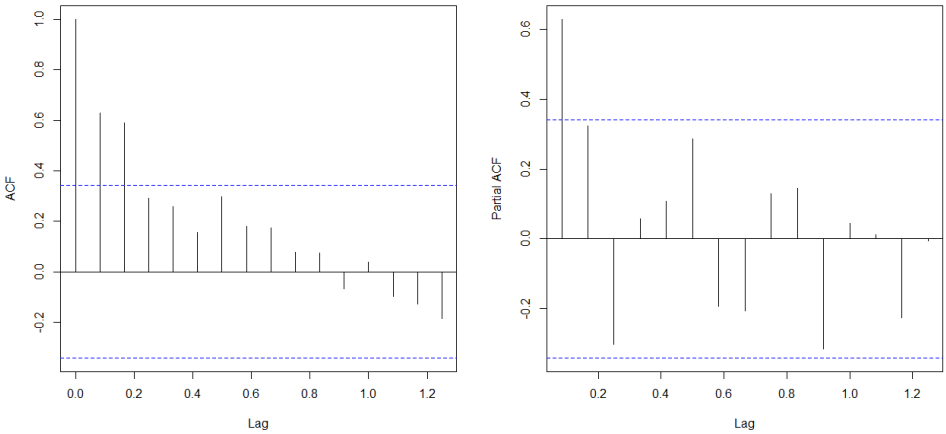
Fonte: autor (2022)

Figura 14– Série do volume de cerveja



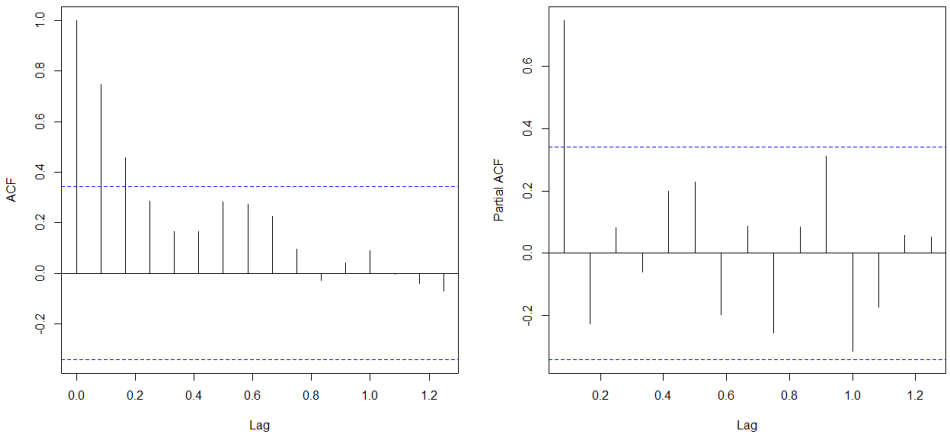
Fonte: autor (2022)

Figura 15 – Série do volume de NAB



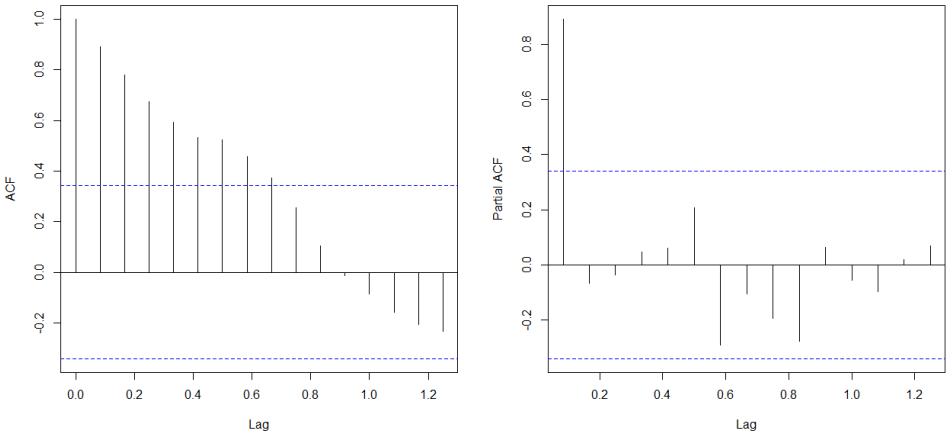
Fonte: autor (2022)

Figura 16 – Série do volume do ASR



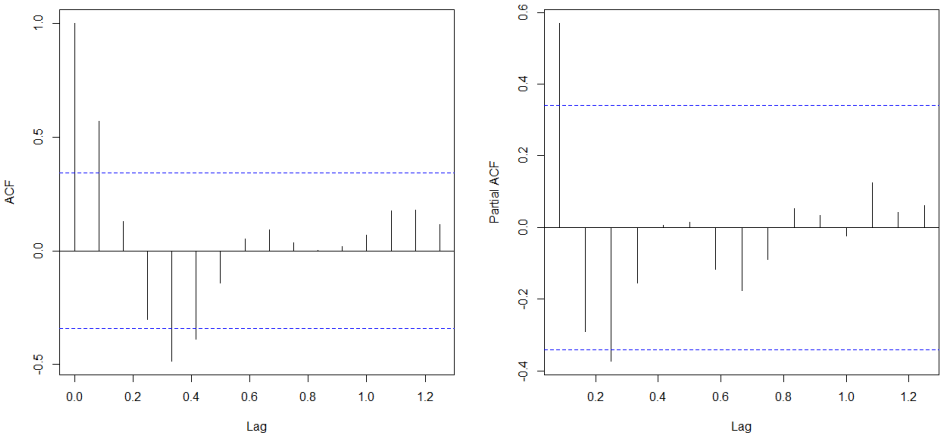
Fonte: autor (2022)

Figura 17 – Série do volume da Central de Bebidas



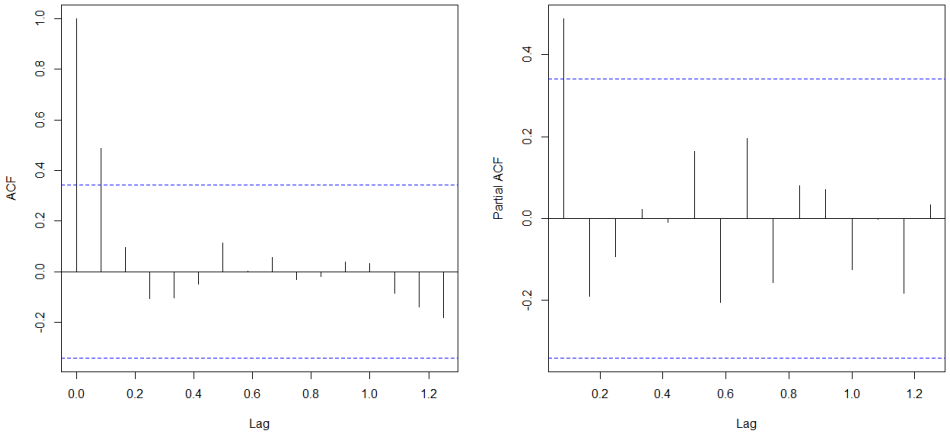
Fonte: autor (2022)

Figura 18 – Série do volume do FRIO



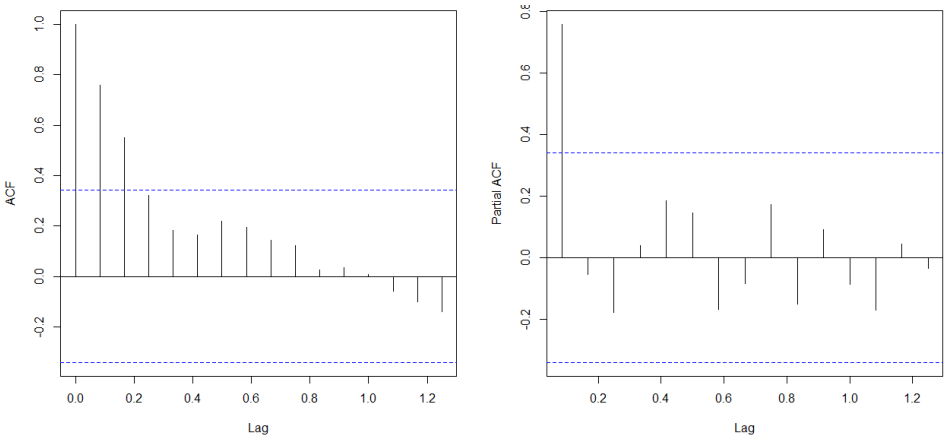
Fonte: autor (2022)

Figura 19 – Série do volume do SUB



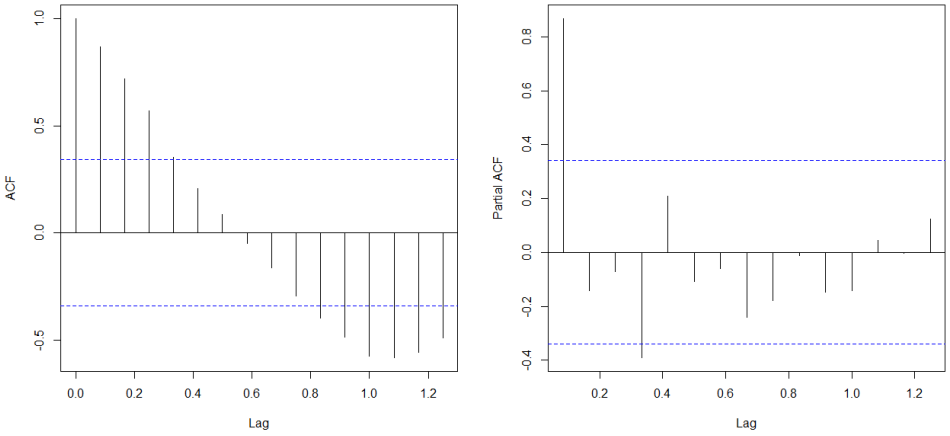
Fonte: autor (2022)

Figura 20 – Série do volume do TRAD



Fonte: autor (2022)

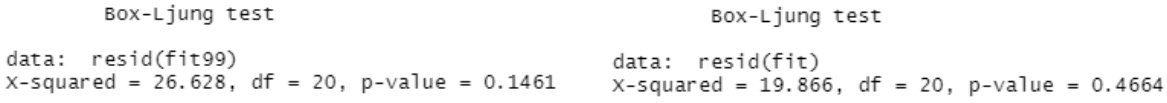
Figura 21 – Série do volume do VIP



Fonte: autor (2022)

APÊNDICE B – Testes de Ljung-Box

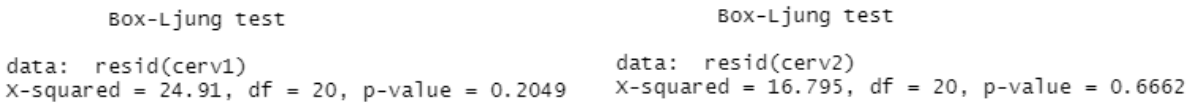
Figura 22 – Série do volume total



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

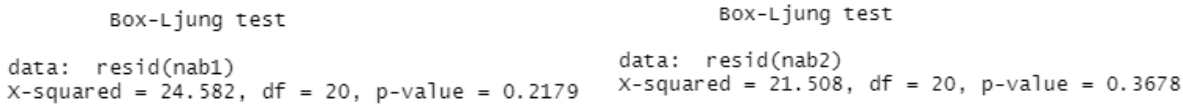
Figura 23 – Série do volume de cerveja



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

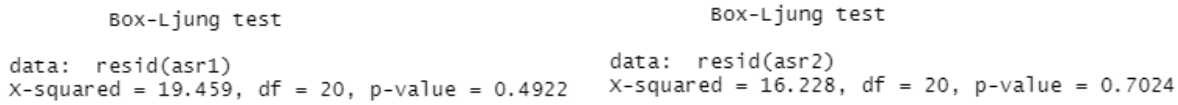
Figura 24 – Série do volume de NAB



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

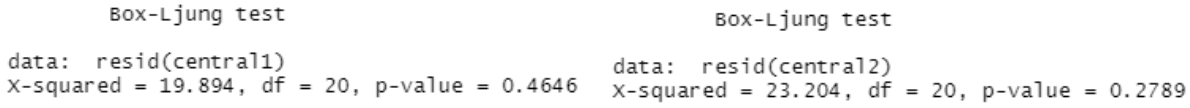
Figura 25 – Série do volume do ASR



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

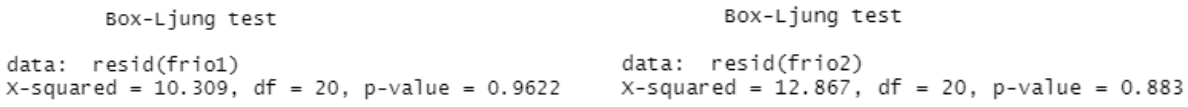
Figura 26 – Série do volume da central de bebidas



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

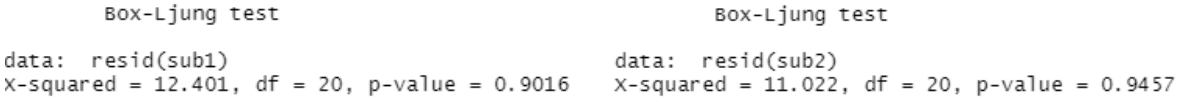
Figura 27 – Série do volume do FRIO



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

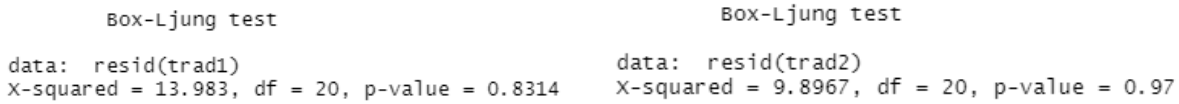
Figura 28 – Série do volume do SUB



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

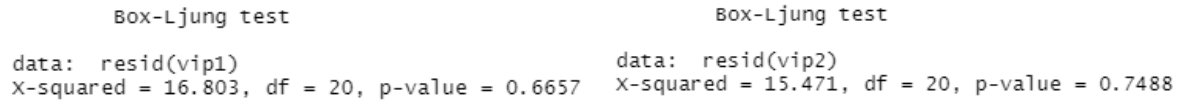
Figura 29 – Série do volume do TRAD



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

Figura 30 – Série do volume do VIP



(a) (b)

(a) Modelo de HW. (b) Modelo ARIMA. Fonte: autor (2022)

APÊNDICE C – Estatísticas descritivas

Tabela 5 – Estatísticas descritivas das séries de dados

	Min.	**1st Qu.**	**Mediana**	**Média**	**3rd Qu.**	**Máx.**
Total	13596	21160	28584	27071	31311	44454
Cerveja	7976	12645	17766	17237	20508	28167
NAB	5062	8160	9855	9834	11173	17001
ASR	5708	9050	12056	11592	13493	19746
Central	0	1078	4942	4297	6340	8524
FRIO	1595	2338	2707	2784	2984	4702
SUB	1749	2313	3094	3007	3514	4459
TRAD	2074	3596	4649	4789	6017	8188
VIP	81	357	547	602	825	1445

Fonte: autor (2022)

APÊNDICE D – Parâmetros dos modelos

Tabela 6 – Parâmetros do modelo HW

	Holt-Winters								
	Total	**Cerveja**	**NAB**	**ASR**	**Central**	**FRIO**	**SUB**	**TRAD**	**VIP**
α	1	0,97	0,86	0,63	1	1	0,36	0,66	1
β	0	0	0	0	0	0	0,20	0	0
γ	0	0	0,40	0	1	0,20	0	0	0

Fonte: autor (2022)

Tabela 7 – Parâmetros do modelo ARIMA

	ARIMA								
	Total	**Cerveja**	**NAB**	**ASR**	**Central**	**FRIO**	**SUB**	**TRAD**	**VIP**
AR_1	1,07	0,92	0	0,86	0,14	0,47	0,79	0	0,90
AR_2	-0,18	0	0	0	0	0	0	0	0
AR_3	-0,02	0	0	0	0	0	0	0	0
MA_1	0	0	-0,30	0	0	0,26	0	0	0
SAR_1	0	-0,27	-0,62	0	0	-0,46	-0,79	0	0

Fonte: autor (2022)

Tabela 8 – Modelos ARIMA utilizados

	ARIMA
Total	(3,0,0)(0,1,0)[12]
Cerveja	(1,0,0)(1,1,0)[12]
NAB	(0,1,1)(1,1,0)[12]
ASR	(1,0,0)(0,1,0)[12]
Central	(1,1,0)(0,1,0)[12]
FRIO	(1,0,1)(1,1,0)[12]
SUB	(1,0,0)(1,1,0)[12]
TRAD	(0,1,0)(0,1,0)[12]
VIP	(1,0,0)(0,1,0)[12]

Tabela 9 – Parâmetros do modelo SVM

	SVM								
	Total	**Cerveja**	**NAB**	**ASR**	**Central**	**FRIO**	**SUB**	**TRAD**	**VIP**
$cost$	50	1000	100	100	10	500	10	10	1000
ϵ	0,50	0,80	0,10	0,30	0,50	0,80	0,05	1	0,80
γ	0	0	0	0	0	0,20	0	0	0

Fonte: autor (2022)