

MODELO DE SÉRIES TEMPORAIS APLICADO A PREVISÃO DE ENTRADA DE PEDIDOS DE ESTRUTURAS DE SISTEMAS DE ARMAZENAGEM

Ronaldo de Oliveira

USP

<http://lattes.cnpq.br/5434631077694977>

E-mail: ronaldo.de.oliveira@outlook.com

Hygor Santiago

UNICAMP - Universidade Estadual de Campinas

<http://lattes.cnpq.br/8858059011756333>

<https://orcid.org/0000-0002-4835-5498>

E-mail: hsantiagolara@gmail.com

DOI-Geral: <http://dx.doi.org/10.47538/RA-2023.V2N4>

DOI-Individual: <http://dx.doi.org/10.47538/RA-2023.V2N4-07>

RESUMO: A grande concorrência entre empresas em busca de inovações e melhorias se torna crescente atualmente, e é natural que as empresas busquem melhorias em processos, capacitação de pessoas e invistam em tecnologia para garantir um diferencial competitivo e melhor participação no mercado (Kagermann; Wahlster, 2013). O presente estudo proporciona uma análise de 12 meses com baixa representatividade da curva real, já que os dados não se comportam conforme o que foi previsto, um dos motivos é que o produto é bem específico e não é tratado como algo sazonal, que há tendência, ou seja, parâmetros que são possíveis prever. Já que, a concepção de uma estrutura de armazenagem são projetos que levam anos para serem realizados o seu fechamento. Como trabalhos futuros, é importante a análise mais específica de uma região com parâmetros que influenciam na tradução de um dado mais fiel, como relacionar a influenciado marketing na venda e concepção de sistemas de armazenagem, ou como as plataformas auxiliam na vendas desses projetos e seus impactos.

PALAVRAS-CHAVE: Modelo de séries temporais. Estruturas de sistemas. Armazenagem.

TIME SERIES MODEL APPLIED TO FORECASTING ORDER ENTRY FOR STORAGE SYSTEMS STRUCTURES

ABSTRACT: The great competition between companies in search of innovations and improvements is growing nowadays, and it is natural for companies to seek improvements in processes, people training and invest in technology to guarantee a competitive advantage and better market share (Kagermann; Wahlster, 2013). The present study provides a 12-month analysis with low representation of the real curve, since the data does not behave as predicted, one of the reasons is that the product is very specific and is not treated as something seasonal, there is a tendency, that is, parameters that can be predicted. Since, the design of a storage structure are projects that take years to complete. As future work, it is important to perform a more specific analysis of a region with parameters that influence the translation of more accurate data, such as relating the influence of marketing on sales and the design of storage systems, or how platforms help in the sales of these projects. and its impacts.

KEYWORDS: Time series model. Systems structures. Storage.

INTRODUÇÃO

A grande concorrência entre empresas em busca de inovações e melhorias se torna crescente atualmente, e é natural que as empresas busquem melhorias em processos, capacitação de pessoas e invistam em tecnologia para garantir um diferencial competitivo e melhor participação no mercado (Kagermann; Wahlster, 2013).

Além disso, a venda é absolutamente essencial para a atividade de uma empresa ou organização. Já que, são pelas vendas que ocorrem o fluxo de recursos que financiam as despesas correntes, os investimentos e a sustentação da própria rentabilidade, conforme abordado por Angelo et al. (2010).

Aires et al. (2019) complementa citando que a previsão de demanda é fundamental e muito importante para a empresa/organização, já que auxilia no aperfeiçoamento da cadeia de suprimentos e contribui para a tomada de decisões em relação a: capacidade produtiva, disponibilidade de mão de obra, campanhas promocionais etc.

O governo de uma cidade, por exemplo, deverá entender qual é o crescimento populacional. Após isso, poderá tomar decisões para realizar o investimento e ampliação em sua infraestrutura básica, como redes de água, luz e esgoto. Da mesma maneira, uma empresa também deve entender e antecipar a demanda de seus produtos, para agendar a sua produção e manter seu estoque em um nível adequado, com isso, entender os principais fatores que influenciam em sua demanda presente, para prever suas decisões adequadas a cada cenário (Passari, 2010).

Angelo et al. (2010) cita que:

Se a previsão é uma função presente em todas as organizações, a avaliação das vendas futuras talvez seja uma das mais importantes e frequentes atividades nesse campo. Por essa razão, a literatura que trata do assunto é rica e extensa. Muitas e variadas técnicas vêm sendo aplicadas em muitos exemplos e casos relacionados aos mais diversos segmentos dos mercados de bens e serviços.

As previsões a fim de contribuir para realizar as predições eficientes são utilizadas através de aprendizado de máquinas ou “*Machine Learning*”. Estas, são efetuadas através de algoritmos que utilizam como base o histórico de vendas passadas, por exemplo, para prever a quantidade de estoque, o custo de aço, produção. Além das previsões com dados internos das organizações é possível identificar comportamentos que influenciam no

comportamento do consumidor, e isso é que torna a previsão por meio de aprendizagem de máquina um diferencial entre outras ferramentas existentes, segundo Pinto et al. (2019).

Sendo assim, é possível analisar com base nos dados anteriores para que uma organização esteja em crescente evolução, é necessário estar alinhado as estratégias de mercado. E para isso é necessário entregar ferramentas que disponibilizam um bom planejamento estratégico que inicia nas vendas. Por isso, o intuito do presente trabalho é desenvolver um modelo preditivo de vendas com abordagem em “*Machine Learning*”, entendendo os fatores interno e externos que influenciam nas vendas de sistemas de armazenagem.

OBJETIVO

Gerar um modelo preditivo balizador de preço de venda que resulta em reais por quilograma na venda de estruturas de sistema de armazenagem, analisando a recorrência das vendas passadas do preço por quilograma de determinados clientes por mês, utilizando faturamento, custos do aço e índices de aumento do aço por região para obtenção de um possível preço de reais por quilograma que possa ser mensurado para futuras vendas.

Os objetivos específicos do presente trabalho são:

- relacionar dados internos e externos a organização;
- obter uma previsão de reais por quilograma de estrutura de armazenagem por mês para futuras negociações;
- proporcionar uma previsão para a compra de bobinas de aço baseada no fechamento de futuras propostas, já que o modelo preditivo obterá os preços de reais por quilograma;
- treinar o modelo preditivo e relacionar com outros modelos preditivos para entender qual será o melhor modelo de aprendizado de máquina para a infraestrutura da empresa no cenário atual;
- Obter os melhores índices que compõem o modelo preditivo.

ANÁLISE DE SÉRIES TEMPORAIS

Segundo Morettin e Tolo (2006), uma série temporal é qualquer conjunto de informações que são ordenadas por um determinado tempo. A série temporal gera interesses como investigar o mecanismo gerador da série temporal, fazer previsões de valores futuros, descrever os comportamentos da série ou procurar periodicidade relevantes nos dados. Sendo assim, a previsão não constitui um fim, mas apenas um meio de fornecer informações para uma consequente tomada de decisões.

Fischer (1982), cita que na análise clássica, costuma classificar os movimentos em quatro tipos básicos de variações:

a) Tendência: caracterizado por aquele movimento regular e contínuo de longo prazo, podendo ser ascendente e descendente que mantêm uma variação;

b) Sazonalidade: Variações que ocorrem com certa regularidade ao longo de um curto período de tempo. Então, aquelas séries que apresentam variações similares de um espaço de tempo a outro, caracterizando-se por mostrarem período de sazonalidade;

c) Variações cíclicas: são flutuações de longo prazo em torno da curva de tendência;

d) Variações aleatórias (ou residuais): gerado por movimentos esporádicos, ocasionados por eventos aleatórios imprevisíveis.

CLASSIFICAÇÃO DAS SÉRIES TEMPORAIS

As séries temporais podem ser classificadas em dois grupos em séries contínuas e em séries discretas. A primeira segue uma quantidade de observações e foram realizadas continuamente em um determinado período de tempo, já as observações discretas são geralmente espaçadas em uma determinada faixa e tempo (CHATFIELD, 1996).

Sendo assim, quando é citado sobre séries temporais é importante salientar que as observações em geral, são consideradas dependentes, e nesse nível de análises pode-se dizer que é possível realizar uma série temporal com valores passados para prever valores futuro e ao final é possível determinar se a série é classificada como determinística (valores exatos de previsão) ou é classificada como estocástica (valores parcialmente previstos) (CHATFIELD, 1996).

OBJETIVO DAS SÉRIES TEMPORAIS

O objetivo da série temporal é criar uma previsão que consiga através de uma função que utilize dados passados em um determinado momento do tempo para estimar valores que a série vai assumir que acontecerão, tentando obter o menor desvio quadrado médio possível entre o valor que foi estimado e o valor que foi previsto BOX et al. (2015).

ESTACIONARIEDADE

Supõe-se que uma série temporal seja estacionária, quando ela se desenvolva aleatoriamente ao longo do tempo sob uma média constante, resultando em uma forma que seguem equilíbrio estável (Morettin; Tolo, 2006).

Antes mesmo de aplicar modelos como média móvel ou autorregressivo, é importante que seja realizado o tratamento da série temporal e seja retirado as fontes de variação, transformando essa série temporal em uma série estacionária, essa transformação consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária (Chatfield, 1996).

Devido à dificuldade de encontrar se as médias e a variância são constantes e finitos dentro de uma série temporal (ruído branco), utiliza-se de ferramentas para identificar se esse conjunto é estacionário ou não é estacionário, conforme afirma Mackinnon(2010). Autocorrelação, autocorrelação parcial e o teste *Augmented Dickey-Fuller* (ADF).

AUTOCORRELAÇÃO (ACF OU FAC)

A autocorrelação mede a correlação entre dois valores adjacentes em uma série, em resumo a autocorrelação é medida como autocorrelação de lag ou defasagem. Em outras palavras significa analisar o impacto que as defasagens podem causar no seu valor atual, no qual representa as médias móveis (Guidolin; Pedio, 2018).

$$P_h = \frac{\sum_{t=h+1}^N (X_t - Y)(X_{t-h} - Y)}{\sum_{t=h+1}^N (X_t - Y)^2} \quad (1)$$

$$Y = \frac{1}{N} \sum_{t=1}^N X_t \quad (2)$$

Onde:

P_h = Autocorrelação;

X_t = variável analisada;

Y = média da amostra;

X_{t-h} = variável analisada em defasagem no instante h.

AUTOCORRELAÇÃO PARCIAL (PACF OU FACP)

São úteis para a identificação de modelo aos dados observados, ou seja, calculado as estimativas das FAC, é possível que seja encontrado as verdadeiras FAC desconhecidas, nessa etapa é necessário que seja escolhido modelos que descrevam o processo estocástico. A FACP representa as ordens autorregressiva, conforme citado por Morettin e Toloí (2006).

$$B_h = \text{Corr}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h+1}) \quad (3)$$

Onde:

B_h = Autocorrelação Parcial;

X_t = variável analisada;

X_{t-h} = variável analisada em defasagem no instante h.

AUGMENTED DICKEY-FULLER (ADF)

Na presença de autocorrelação, a diferenciação de primeira ordem, pode ser representada por uma regressão linear de tempos anteriores e resultará até um atraso (lag)

de índice m que partirá dos diferenciais de primeira ordem. A série não será estacionária se a raiz unitária e a hipótese nula ser aceita. Caso contrário, a série será interpretada como estacionária e a hipótese nula é rejeitada (Pal; Prakash, 2017).

MODELO AUTORREGRESSIVO - AR(Q)

Uma série temporal (X_t) é classificada com ordem p , se ocorrer a soma linear ponderada dos valores p anteriores, somado a mais um componente de aleatoriedade (CHATFIELD, 2000).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t \quad (4)$$

Onde:

X_t = média autoregressivo;

ϕ_p = coeficiente de regressão do modelo;

X_{t-p} = valores no instante t somando a um termo de característica aleatória p ;

Z_t = termo característica aleatória. Resultará em um ruído branco caso a modelagem for perfeita em um instante t .

O modelo de autorregressão utiliza das observações anteriores como *input* para uma equação de regressão e assim, consegue realizar a previsão do valor na próxima observação. O processo pode gerar previsões em diversas problemáticas, apesar de simples. Conforme afirma Brownlee (2020), o nome autorregressivo surgiu em função da equação utilizar dados da mesma variável de entrada comparando em etapas de tempo anteriores.

$$X_{t+p} = b_0 + b_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t \quad (2)$$

Onde:

X_{t+p} = Quantidade de inputs de entrada;

b_0 = Coeficientes encontrados através da otimização do modelo de treino;

p = Quantidade de regressões realizadas;

MODELO MÉDIA MÓVEL - MA(D)

Dentro da Série temporal (X_t), pode-se definir um processo como média móvel de ordem q , se for realizado uma soma linear ponderada dos últimos q choques aleatórios, apresentado na equação (3). Já o Z_t é um processo aleatório com médio zero e variância constante.

$$X_t = Z_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_q Z_{t-q} \quad (3)$$

Onde:

X_t = Série temporal por média móvel;

Z_t = Termo aleatório com média zero e variância constante;

q = Quantidade de choques aleatórios que foram considerados no modelo;

ϕ_q = Coeficientes que ponderam os choques;

Z_{t-q} = Choques aleatórios que foram considerados no modelo.

Sendo assim, o processo de média móvel em uma série temporal consegue trabalhar com qualquer processo estacionário, resultando na soma de dois tipos de processos (CHATFIELD, 2000).

O valor da média móvel pode ser utilizado para gerar previsões diretamente e o seu processo é um modelo que só é aplicável quando novas observações estão disponíveis, ou seja, é realizado uma iteração entre a primeira previsão, em seguida coleta a nova observação, atualiza o modelo e realiza, novamente, a previsão seguinte com o dado da previsão. Já o processo de suavização por média móvel resulta na remoção do ruído entre os dados (observações) e na exposição do melhor sinal (BROWNLEE, 2020).

MODELO AUTORREGRESSIVO INTEGRADO A MÉDIA MÓVEL - ARIMA

A classe do modelo ARIMA é uma importante ferramenta de previsão que foi inicialmente realizada por Box e Jenkins (CHATFIELD, 1996). O Modelo ARIMA

(p,d,q) consiste em ajustar modelos auto-regressivos integrados de médias móveis (Morettin; Toloi, 2018). O modelo é dividido em:

- AR(q): componentes autorregressivos que utilizam da relação entre a dependência entre a observação em um determinado período;
- Integrada (d): as diferenças para tornar uma série estacionária;
- MA (q): médias móveis que utilizam a dependência entre uma observação e um erroresidual.

Segundo Morettin e Toloi (2006), o modelo ARIMA segue os seguintes ciclos iterativos:

- a) Uma classe geral de modelos é considerada para a análise (especificação);
- b) Há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- c) a seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- d) finalmente, há a verificação ou diagnóstico do modelo ajustado, através de uma análise de resíduos, para se saber se este é adequado para os fins em vista (previsão, por exemplo).

Caso o modelo determinado, não esteja de acordo, o ciclo é repetido, voltando-se à fase de identificação, pode ser que diferentes pesquisadores identifiquem diferentes modelos para a mesma série temporal. É um modelo muito utilizada quando relacionado com previsão de séries temporais, pois compões de diferentes estruturas temporais em uma série e resulta em um método simples para realizar as previsões (BROWNLEE, 2020).

A obtenção dos parâmetros p, d e q para o modelo ARIMA podem ser visualizados nos gráficos de função de autocorrelação de amostra (ACF) e em função de autocorrelação parcial(PACF), conforme afirma Hyndman e Athanasopoulos (2018).

MODELO AUTORREGRESSIVO INTEGRADO A MÉDIA MÓVEL – SARIMA (PYTHON AUTO-ARIMA)

O que diferencia o modelo ARIMA do modelo SARIMA é que no primeiro

modelo, apenas variáveis de autorregressivo, integrado e média móvel representante pelas letras: p, d e q. Já no SARIMA, S que significa sazonal, inclui além das citadas anteriormente, as variáveis sazonais de autorregressivo sazonal (P), Integrado sazonal (D) e média móvel sazonais (Q) e período sazonal (S) que indica o tempo (SEABOLD; PERKTOLD, 2010).

METODOLOGIA

TRATAMENTO DOS DADOS

Os dados utilizados no presente trabalho foram extraídos de uma empresa em que constitui em seu portfólio de venda produtos de sistema de armazenagem. Vale ressaltar que para manter os dados da empresa seguro foi utilizada uma base amostral para não revelar o faturamento real da empresa por linha de produto.

Na Tabela 01 está disponível uma amostra de como está disponibilizado os dados para análise que gerará o modelo de previsão de séries temporais. Vale lembrar que a base já estava tratada e a único processo de tratamento de dados foi converter o peso, o valor total pedido e adatar em seus respectivos tipos de dados. Além disso, criou-se o campo Ano e Mês para realizar o agrupamento dos dados, visto que a data estava separada por dia.

Tabela 1. Amostra dos dados originais que compõe o estudo e os nomes das respectivas colunas

Desc.	Grupo Gerencial Produto	Data	Tipo Frete	Região	Uf	peso	valor total pedido
Tipo de dados	object	Datetime64 [ns]	object	object	object	float64	float64
Dados	PORTA PALLET	01/09/2020	CIF	Sudeste	ES	385	812
	PORTA PALLET	06-08-2021	CIF	Sudeste	SP	399.131	812.429
	DRIVE IN	28-09-2020	CIF	Sudeste	SP	641.256	8.083.582
	SERVICOS	19-01-2023	CIF	Sudeste	SP	0	7.534.851
...							

Fonte: Do autor (2023).

REGRA DE NEGÓCIO

As estruturas de armazenagem são divididas em vários grupos, tais como: Porta Pallet, Drive-in, Push Back, Dinâmico etc. Cada estrutura tem um conceito específico de funcionamento. Sendo assim, o presente estudo utilizará os dados históricos da estrutura de Porta Paleta que tem a maior quantidade de dados histórico presente na base, conforme Tabela 02. Já que o Porta Pallet foi selecionado para o presente estudo, deve-se levar em consideração a região de análise, já que cada região tem um comportamento de compra diferente para cada estrutura.

Sendo assim, agrupando o grupo gerencial do produto e agrupando-o por região é possível concluir que o total de registros utilizados no modelo preditivo de vendas será de 884 registros, conforme Tabela 03, se estendendo de 01/01/2020 até 01/04/2023.

Tabela 02. Quantidade de linhas por linha de produto.

Grupo Gerencial Produto	Nº de linhas
AUTOMACAO	32
AUTOPORTANTE	25
BASES MOVEIS	3
BOBINAS	1
CANTILEVER	24
CARRO SATELITE	26
DINAMICO	122
DRIVE IN	406
DRIVE THROUG	10
FLOW RACK	74
MEZANINO	30
NÃO INFORMADO	60
OUTROS	368
PASSARELA	63
PORTA BOBINA	12
PORTA PALLET	2410
PUSH BACK	52
SERVICOS	594
TRANS NAO MOTOR	1
TRANSELEVADORES	8
TUBOS	1
UNIVERSAL	243
Total	4565

Fonte: Do autor (2023).

Tabela 03. Quantidade de linhas por grupo gerencial do produto e por região.

Grupo	Gerencial	Região	Nº de linhas
PORTA PALLET		Centro-Oeste	128
PORTA PALLET		Exterior	59
PORTA PALLET		Nordeste	311
PORTA PALLET		Norte	69
PORTA PALLET		Sudeste	884
PORTA PALLET		Sul	959
Total Geral			2410

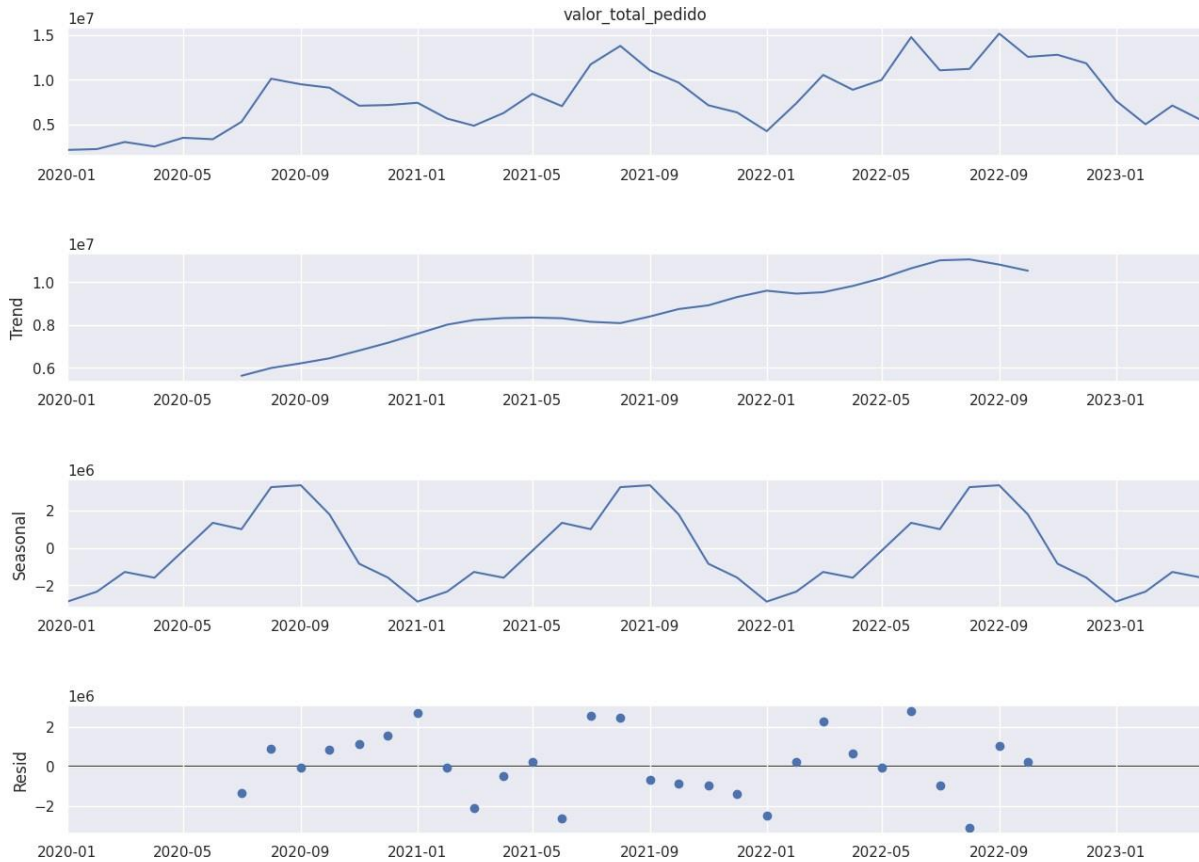
Fonte: Do autor (2023).

Na Figura 01, pode-se analisar que há picos de valores recorrentes a cada 12 meses. A série não se comporta em uma tendência de compra crescente, porque geralmente são clientes novos que adquirem estruturas ou podem até ser os mesmos clientes, porém para regiões diferentes. Cada pedido dos 884 para a região sudeste é um projeto específico ou reposição para projetos que já foram comprados, estruturas de armazenagem não se comportam que nem o segmento de varejo em função da recorrência de compra e de projetos grandes que podem ser adquiridos em um único projeto e que ocorre a demanda de meses de conclusão da realização dos pedidos (orçamento e projeto), fechamento do pedido, fabricação das peças, faturamento das peças e montagem das estruturas “in loco”.

Analisando a decomposição da série temporal na Figura 01, a tendência apresenta um aumento ao longo dos anos separada por três retas crescentes e duas retas com um leve decaimento, conforme Figura 02.

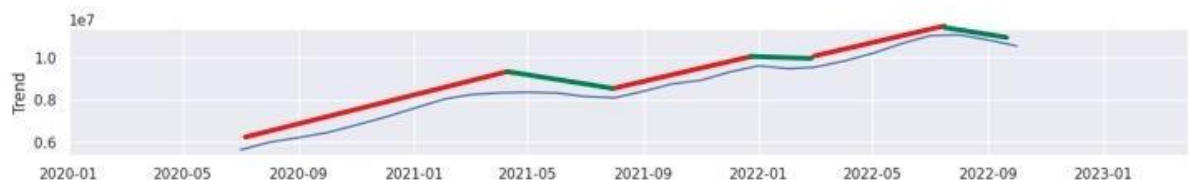
Retirando o ruído, pode-se perceber que com os pontos da sazonalidade e das tendências quando somados geram uma possível função que resulta em uma equação. Essa função sem os ruídos poderia gerar um possível gráfico do valor total do pedido.

Figura 01. Decomposição da Série Temporal – Valor total do pedido em função do tempo



Fonte: Do autor (2023)

Figura 02. Tendência oscilando em crescimento (reta vermelha) e decrescimento (reta verde)



Fonte: Do autor (2023)

É fundamental que seja analisado se a série é estacionário, caso não seja é necessário torná-la estacionária. Então utilizou-se a função do Dickey-Fuller para verificar se há ou não Estacionariedade na série. Como o p-value é menor que 0,05, então a função rejeita o h_0 e a série pode ser caracterizada como estacionária.

Além disso, o valor crítico da série que representa o ADF “Statistic”, resultou em aproximadamente -3,3782, analisando-se os valores de 1%, 5% e 10% de confiança, é possível observar que o valor crítico de 1% é maior que o valor crítico da série, contudo

menor que o nível de confiança de 5% de -2,949. Em outras palavras, a série fornece um nível de confiança de 95%, conforme apresenta a Tabela 04.

Tabela 04. Resultado da Estacionariedade dos dados

ADF Statistic		-3,3782
p-value		0,01173
Critica Values	1%	-3,633
	5%	-2,949
	10%	-2,613
É estacionaria!		

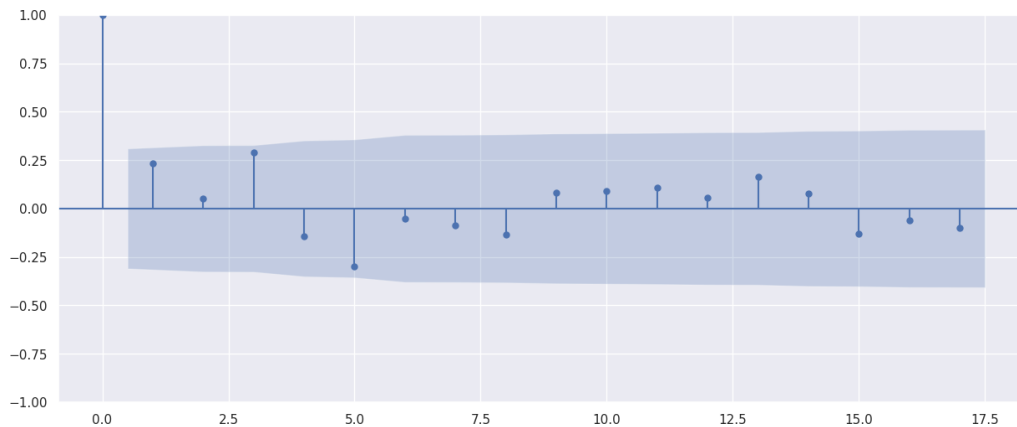
Fonte: Do autor (2023)

Após determinar a Estacionariedade dos dados é necessário encontrar as ligações entre as observações passadas e os presentes, ou seja, aplicar o método de autocorrelação, no qual, apresentará a correlação entre a série e ela mesma, por esse motivo é designado autocorrelação, porém ela defasada.

Vale ressaltar que a correlação varia entre 0 e 1 na ordenada e na abscissa os valores de defasagem. Como é possível verificar na Figura 03, a área hachurada representa o nível de significância da série temporal, e os valores são significativamente diferentes de zero o que resulta em uma existência de autocorrelação com os dados passados.

Quanto mais os valores próximos de 1 mais autocorrelação com os dados com uma defasagem específica, mais autocorrelação existe na série, na Figura 04, pode-se notar que há uma fraca autocorrelação entre os valores em defasagem, ao não ser o lag 0, que é 1 em função do primeiro valor ser relacionar com ele mesmo.

Figura 03. Autocorrelação

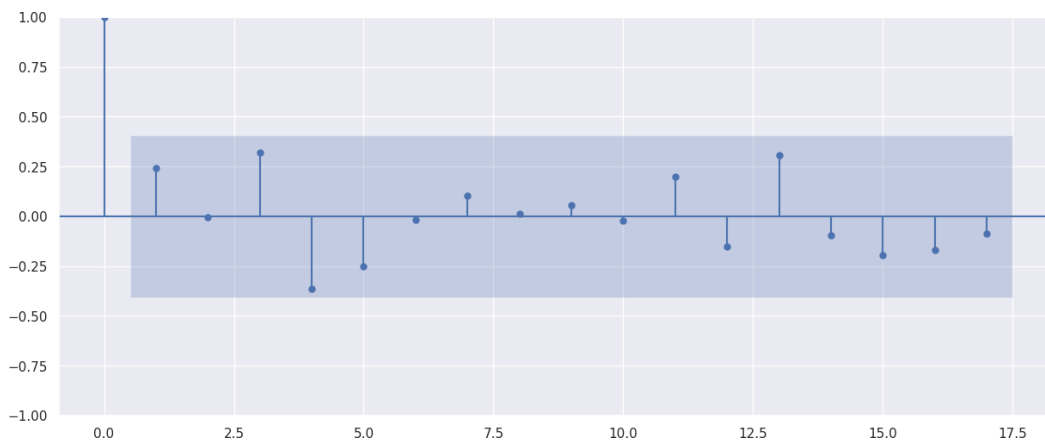


Fonte: Do autor (2023).

Analisando a autocorrelação que indica as influências diretas ou indiretas da série temporal, o próximo passo é analisar a autocorrelação parcial que considera somente a correlação direta entre a série a sua versão defasada.

Conclui-se que a autocorrelação, também há uma fraca autocorrelação parcial, porém seu comportamento, conforme Figura 04, é totalmente diferente do comportamento da Figura 04, onde é apresentado as influências diretas e indiretas da série. A autocorrelação parcial apresenta sentidos semelhantes no início dos lags, porém à medida que os lags se aproximam de 17,5 os valores da autocorrelação se modificam o sinal.

Figura 04. Autocorrelação Parcial



Fonte: Do autor (2023)

Com as análises realizadas da Estacionariedade e autocorrelação da série, o passo seguinte será a identificação dos parâmetros do ARIMA. Para isso será necessário utilização do conceito de Box-Jenkins, porém o Python disponibiliza de uma função auto-ARIMA. Em resumo são estimados os parâmetros de entrada e a rotina do Python retorna o melhor modelo em p,d,q,P,D e Q da série. Vale ressaltar que mesmo que a série foi identificada como estacionário, optou-se em utilizar o modelo ARIMA para realizar a previsão.

Conforme Tabela 5 é possível verificar claramente que o modelo foi estimado para 12 meses e que o modelo é claramente estacionário pelos parâmetros de entrada p,d e q resultarem em 0. A base de treino do modelo se estende do dia 01/01/2020 até 01/01/2022 e a base de teste utilizada no modelo é todos os dados maiores que 01/01/2023.

Tabela 5. Modelo Auto-ARIMA

Descrição do modelo Auto Arima	Coefficiente AIC
ARIMA(0,0,0)(0,1,1)[12]	AIC=967.681

Fonte: Do autor (2023).

Realizado os parâmetros para o modelo ARIMA (0,0,0) (0,1,1) [12], gerou-se uma previsão com dados com base no período da base de teste resultando valores de 01/01/2022 até 01/04/2023. É possível analisar que as médias móveis e o autorregressivo são dois parâmetros que podem ser analisados nos gráficos de fac e facp, como analisado anteriormente, esses parâmetros não têm autocorrelação e quando eu realizo o processo do ARIMA, os parâmetros resultam em 0, respectivamente. Os resultados encontrados do modelo real com o previsto podem ser visualizados na Figura 01.

Figura 05 – Valor real (treino) e valor previsto



Fonte: Do autor (2023)

Na Tabela 6 estão disponíveis os valores dos valores reais comparados com os valores previstos encontrados no modelo anterior, juntamente com a diferença entre eles, resultando em uma base percentual. Quanto mais próximo de zero os percentuais melhores são os resultados encontrados entre o valor real e o valor previsto.

Tabela 06. Comparação percentual entre o valor real e o valor previsto

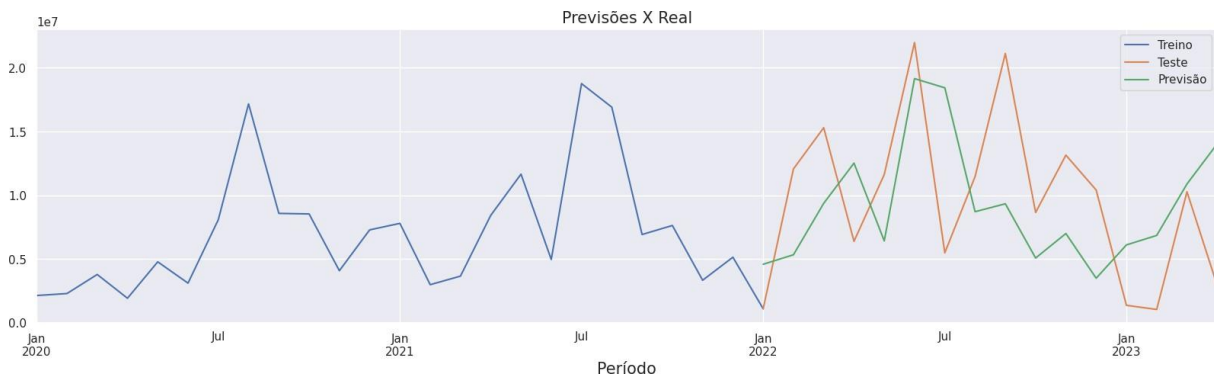
data_agrupada	valor_real	valor_previsto	dif	% dif
01/05/2022	1.092.852	4.585.373	-3.492.521	-76%
01/06/2022	12.066.025	5.330.292	6.735.733	126%
01/07/2022	15.300.644	9.367.109	5.933.535	63%
01/08/2022	6.384.801	12.522.190	-6.137.389	-49%

01/09/2022	11.628.771	6.417.996	5.210.775	81%
01/10/2022	21.982.237	19.151.510	2.830.727	15%
01/11/2022	5.481.328	18.430.590	-12.949.262	-70%
01/12/2022	11.462.899	8.708.860	2.754.039	32%
01/01/2023	21.126.865	9.330.914	11.795.951	126%
01/02/2023	8.648.332	5.073.916	3.574.416	70%
01/03/2023	13.147.581	6.998.761	6.148.820	88%
01/04/2023	10.412.239	3.490.484	6.921.755	198%

Fonte: Do autor (2023)

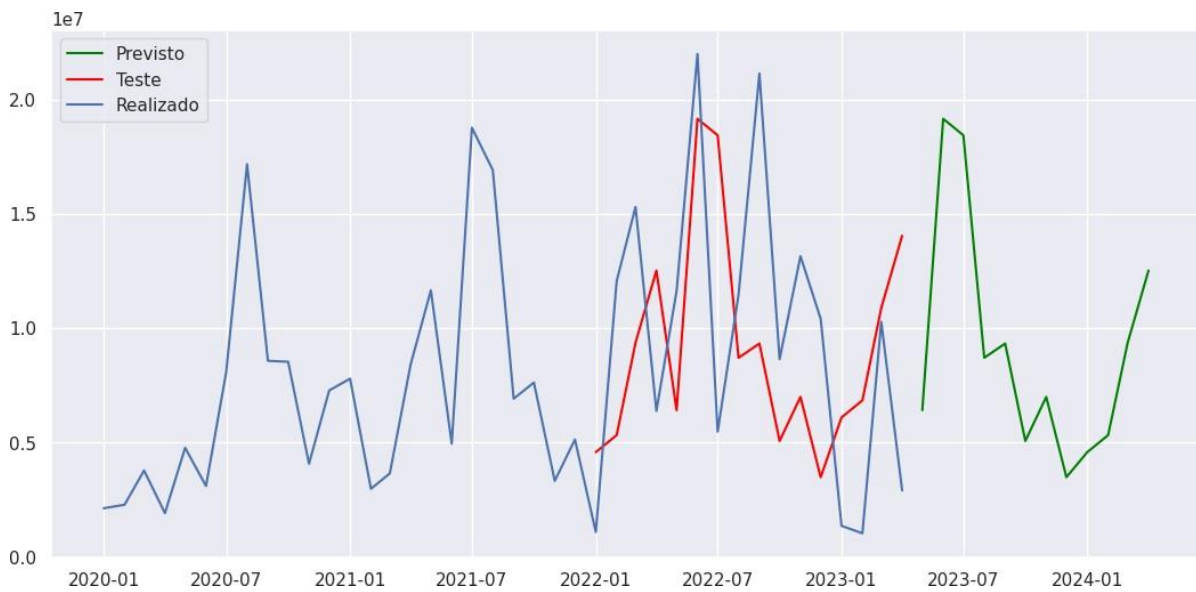
Na Figura 06 e na Figura 07 estão disponíveis as 3 curvas, a primeira em azul representao treino do modelo, a linha em vermelho os dados que foram utilizados para testar o modelo e a previsão que foi o modelo preditivo criado a partir da base de teste. A base de treino servirá para encontrar os parâmetros que estão relacionam os dados passados para tentar prever os dados que estão na base de teste.

Figura 06 – Valores reais e valores de previsão do modelo



Fonte: Do autor (2023)

Figura 07 – Valores da base de teste, o realizado e o previsto.



Fonte: Do autor (2023).

CONCLUSÕES E CONSIDERAÇÕES

Com base na metodologia utilizada no modelo ARIMA combinando com a sazonalidade do auto-arima (Sarima) da linguagem python foi possível analisar que há predominância do ruído branco na série temporal, o que foi possível analisar quando verificamos a autocorrelação e autocorrelação parcial, em quem, todos os lags resultaram na faixa em que determina nenhuma autocorrelação. Além disso, foi possível concluir que a série é estacionária e quando realizamos o método ARIMA é comprovado em que o resultado dos parâmetros p , d e q resultam em 0.

Além disso a estacionariedade da série temporal com um nível de confiança de 95%, resultando em um p -value de 0,011 que é traduzido na hipótese nula não rejeitada, ou seja, em que a série é estacionária.

O presente estudo proporciona uma análise de 12 meses com baixa representatividade da curva real, já que os dados não se comportam conforme o que foi previsto, um dos motivos é que o produto é bem específico e não é tratado como algo sazonal, que há tendência, ou seja, parâmetros que são possíveis prever. Já que, a concepção de uma estrutura de armazenagem são projetos que levam anos para serem realizados o seu fechamento.

Como trabalhos futuros, é importante a análise mais específica de uma região com parâmetros que influenciam na tradução de um dado mais fiel, como relacionar a influenciado marketing na venda venda e concepção de sistemas de armazenagem, ou como as plataformas auxiliam na vendas desses projetos e seus impactos.

REFERÊNCIAS

ANGELO, C.; ZWICKER, R.; FOUTO, N.; LUPPE, M. **Séries temporais e redes neurais: uma análise comparativa de técnicas na previsão de vendas do varejo brasileiro**. Brazilian Business Review, v.8, n.2, p. 01-21, 2011.

AIRES, C.S.F; ALMEIDA, G.J.A; SILVEIRA, S.O; 2019. **Inteligência artificial na gestão de estoque, X Fateclog Logística 4.0 e a sociedade do conhecimento**, FATEC, Guarulhos, SP, Brasil.

BROWNLEE, J. **Introduction to time series forecasting with python: how to prepare data and develop models to predict the future**. Machine Learning Mastery, 2020.

BLACKWELL, R. D.; ENGEL, J. F.; MINIARD, P. W. **Comportamento do consumidor**. Tradução: Eduardo Teixeira Ayrosa (coord.). São Paulo: Cengage Learning, 2008.

CHAPMAN, P; CLINTON, J; KERBER, R; KHABAZA ,T; REINARTZ, T; SHEARER, C; WIRTH, R. **Step-by-step data mining guide**. SPSS inc, 2000.

CHATFIELD, C. **Time-series forecasting**. CRC press, 2000.

GUIDOLIN, M.; PEDIO, M. **Essentials of time series for financial applications**. Academic Press/Elsevier, 2018.

MACKINNON, J. G. **Critical values for cointegration tests**. Queen's Economics Department Working Paper, n. 1227, 2010. Disponível em: <https://www.econstor.eu/bitstream/10419/67744/1/616664753.pdf>. Acesso em: 04 jun. 2023.

AGERMANN, H. WAHLSTER, W. 2013. **Securing the future of Germany manufacturing industry: recommendations for implementing the strategic initiative Industrie 4.0 – final report of the industrie 4.0 working group**. Essen: Forschungsunion; Munich: Acatech, 2013.

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

PASSARI, A.F.L. 2010. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. Dissertação de Mestrado em Administração. Universidade de São Paulo, São Paulo, SP, Brasil.

PINTO, I.B; ZORZO, A; SCHLÜTER, M, R; 2019. **Aplicação do aprendizado de máquinas na previsão de demanda**. ISSN 2357-9684, XII Fateclog gestão da cadeia de

suprimentos no agronegócio: desafios e oportunidades no contexto atual FATEC, Mogi das Cruzes, SP, Brasil.

SEABOLD, S.; PERKTOLD, J. **Statsmodels**: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference, Austin, 2010.

Submissão: junho de 2023. Aceite: setembro de 2023. Publicação: outubro de 2023.