



ConBRepro

XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



IA nas Engenharias

29 nov. a 01
de dezembro 2023

ANÁLISE COMPARATIVA DA APLICAÇÃO DE MÉTODOS QUANTITATIVOS PARA PREVISÃO DE TENDÊNCIAS DE BUSCAS NA INTERNET DE UMA EMPRESA DO RAMO DE BEBIDAS

Matheus Pereira de Andrade

Programa de Pós-Graduação em Engenharia de Produção - UFPR

Mariana Kleina

Programa de Pós-Graduação em Engenharia de Produção - UFPR

Resumo: A Coca-Cola é um dos refrigerantes mais vendidos no mercado de bebidas no mundo e está no mercado desde 1892, a sua companhia é patrocinadora de diversos campeonatos de esportes e outros eventos no mundo e isso faz com que ela sempre fique em evidência, não só pela qualidade do seu produto, mas também pelas grandes campanhas de *marketing* que a empresa realiza, o que faz com que a busca pelo seu nome na internet, aumente cada vez mais. O trabalho desenvolvido nesse artigo visa apresentar uma análise comparativa dos métodos quantitativos de previsão de demanda utilizando os dados extraídos no *Google Trends* para identificar qual deles se ajusta melhor na previsão da palavra-chave “Coca-Cola”, visando escolher aquele que apresentará menor índice de erro. Essa análise é realizada por meio dos *softwares* de apoio Microsoft Excel e R. Essa análise elenca os itens relevantes para a verificação e utilização dos métodos escolhidos. No final, é apresentado o método que pode ser escolhido pela empresa para prever em quais momentos ela será tendência e pode determinar possíveis estratégias de *marketing* ou divulgação para aumentar sua popularidade em determinado período.

Palavras-chave: *Google Trends*, Coca-Cola, Métodos quantitativos, Previsão de demanda.

Comparative analysis of the application of quantitative methods to forecast internet search trends of a beverage company

Abstract: Coca-Cola is one of the best-selling soft drinks on the beverage market in the world and has been on the market since 1892, the company is a sponsor of several sports championships and other events around the world and this makes it always stand out, not only for its quality of your product, but also due to the large marketing campaigns that the company carries out, which causes the search for your name on the internet to increase more and more. The work developed in this article aims to present a comparative analysis of quantitative demand forecasting methods using data extracted from *Google Trends* to identify which one fits best in predicting the keyword “Coca-

Cola”, aiming to choose that will present the lowest error rate. This analysis is carried out using Microsoft Excel and R support software, with the support of equations of the determined methods that best fit, according to what is presented in the literature. This analysis lists the relevant items for verifying and using the chosen methods. In the end, it shows which method can be chosen by the company to predict when it will be a trend and be able to determine possible marketing or advertising strategies to increase your popularity in a given period.

Keywords: *Google Trends*, Coca-Cola, Quantitative Methods, Demand Forecast.

1. Introdução

Segundo Ramsawak *et al.* (2023), o *Google Trends* é um site público que oferece dados com base na frequência com que determinado termo de pesquisa é usado na plataforma de pesquisa do Google INC. As consultas de pesquisas por palavras são definidas em termos ou tópicos. Muitas pessoas acabam realizando a todo momento pesquisas rápidas no *Google* popularizando as pesquisas no principal buscador do mundo na atualidade.

A maior empresa de não alcoólicos do mundo (PEZZOTI, 2022), a Coca-Cola, é uma empresa que tem um grande apelo, não só pela qualidade do seu produto, mas também, por ser uma marca muito forte e não vender apenas pela sua qualidade, mas também pela experiência que ela proporciona. É costumeiro ver as Copas do Mundo e Jogos Olímpicos com patrocínio dessa empresa e indo para um nicho diferente, as campanhas de Natal são aquelas que marcam bastante na memória das pessoas. Isso faz com que a “Coca-Cola” seja um ótimo item de pesquisa para averiguar a sua “popularidade”.

Para encontrar resultados satisfatórios, serão elencados métodos quantitativos de previsão que são viáveis para utilização em uma hipotética previsão e, dentre eles, escolher qual é o melhor, isto é, que apresenta o melhor modelo de previsão possível, para que futuramente, possa ser utilizado pela empresa para prever momentos em que a marca pode estar em baixa nos buscadores e a empresa poder tomar alguma decisão para impulsionar a marca nesses momentos.

Este artigo está dividido em 5 grandes partes, a primeira introduz o leitor ao tema do artigo. A segunda parte, cita o passo-a-passo que foi utilizado para realizar esse artigo, desde a identificação da questão até a conclusão de qual o melhor método para se utilizar nesse caso. A terceira seção traz o referencial teórico acerca de séries temporais, métodos de previsão de demanda e medida de erro de previsão. A quarta seção apresenta os resultados alcançados com a aplicação de dois métodos de previsão à série temporal escolhida, bem como os erros de cada método. A quinta seção conclui o estudo, incluindo sugestão para trabalhos futuros.

2. Metodologia

Os passos que foram realizados para que pudessem ser alcançados resultados coesos para este artigo foram divididos em cinco etapas, conforme Figura 1.

A 1ª etapa é a definição e compreensão do tema, verificação do problema que é retratado por esse tema e se é possível a construção de uma linha de pesquisa coesa, baseada na análise comparativa da aplicação de métodos quantitativos para previsão de buscas futuras sobre o nome Coca-Cola de acordo com o *Google Trends*.

A 2ª etapa é a coleta dos dados na plataforma *Google Trends*, foi necessária a compreensão de como se utiliza e quais dados poderiam ser extraídos dela e posteriormente para armazenamento e utilização dos dados para o estudo. Para realizar esse estudo foi usada a palavra-chave “Coca-Cola” para verificar qual é a tendência de busca pelo nome no Brasil. Os dados são separados semanalmente, com um total de 261

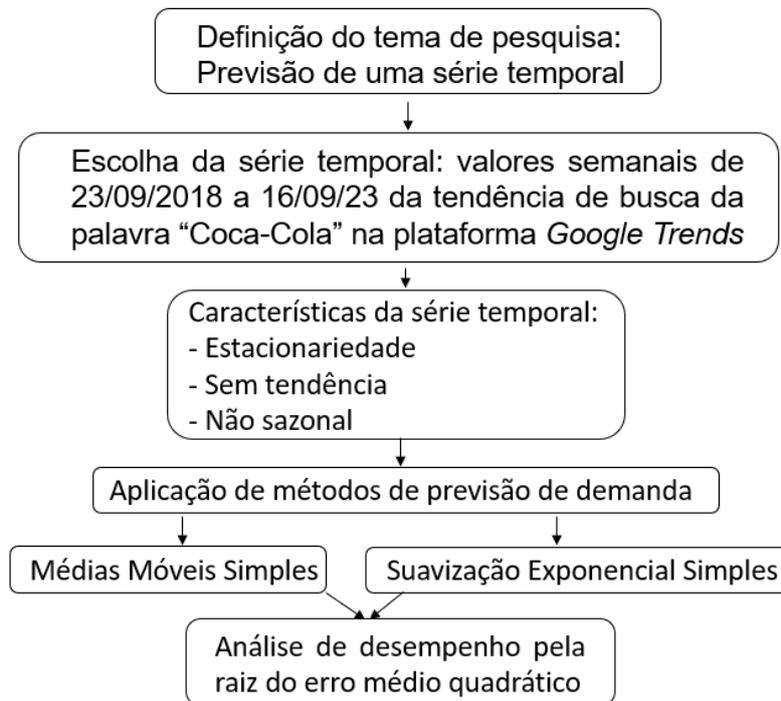
dados para estudo. O período analisado é do dia 23/09/2018 à 16/09/2023, um intervalo de 5 anos.

A 3ª etapa é a análise das características da série temporal obtida. As características observadas da série em questão são: estacionariedade, sem tendência e sem sazonalidade. Estas características foram confirmadas com o auxílio do *software* R.

Na 4ª etapa são escolhidas as ferramentas utilizadas para a obtenção dos resultados, tais como os métodos de previsão de demanda e a medida de erro utilizada. A saber, neste trabalho foram utilizados os métodos de Médias Móveis Simples e Suavização Exponencial Simples para as previsões, e a raiz do erro médio quadrático para comparação de resultados. Para esta etapa, fez uso do *software* Microsoft Excel.

Na 5ª etapa é realizada a aplicação dos métodos quantitativos de previsão e verificado qual é o método mais eficiente para realizar as previsões, de acordo com o cálculo dos erros.

Figura 1 – Etapas da pesquisa



Fonte: Os autores (2023)

3. Referencial Teórico

O referencial teórico foi feito de forma sistêmica e essa sessão foi planejada para exibir as ferramentas utilizadas no estudo.

3.1 Google Trends

O *Google Trends* serve para determinar quais são as tendências de busca no *Google Search*, *Google News* e *Youtube*. Ele serve para identificar o que está como tendência no mundo, o que as pessoas estão seguindo ou não (*GOOGLE, 2023*).

Ramsawak *et al.* (2023) citam que as consultas de pesquisa por palavras são definidas em termos ou tópicos. As consultas com base nos termos de pesquisa incluem correspondências para uma palavra ou frase exata feita na caixa de pesquisa do *Google*. Já as pesquisas por tópicos são mais amplas e incluem termos relacionados que compartilham o mesmo conceito em qualquer idioma.

Segundo Springer *et al.* (2022), os dados gerados pelo *Google Trends* são objeto de estudo para várias áreas. Por exemplo, na área da saúde, para identificar possíveis casos de COVID-19, as pessoas realizam a sua busca de alguns dos sintomas e chegam à busca final sobre a doença.

3.2. Séries Temporais

A análise de séries temporais se aplica nos casos em que há um padrão persistente ou sistemático no comportamento da variável, que é possível de captar por meio de uma representação paramétrica (PINDYCK e RUBENFIELD, 2002). Um dos métodos clássicos citados projetados para uma categoria de dados são os dados com tendência. Segundo Morettin e Toloí (2004), pode-se entender a tendência como o movimento persistente dos dados em uma dada direção.

Segundo Veríssimo *et al.*, (2012), o objetivo da análise de séries temporais é encontrar modelos (estatísticos e/ou matemáticos) que permitem descrever de forma adequada o processo de geração de dados para prever o comportamento futuro das variáveis para um determinado horizonte de planejamento.

3.3 Previsão de Demanda

Segundo Moreira (2008) o papel desempenhado pela previsão é de grande importância em diversas áreas de gestão nas organizações, pois auxilia na visualização de cenários futuros e planejamento de ações, buscando informações sobre vendas de um item ou conjunto deles.

De acordo com Luft (2002), o verbo prever significa ver, saber, examinar antecipadamente, antever, conjecturar, pressupor, fazer suposições e calcular e no final, cita que a palavra previsão é o seu substantivo feminino. Com isso, pode-se assumir que previsão é a estimativa de valores no futuro, seja eles, uma ou várias variáveis.

Uma previsão adequada deve dar suporte a uma decisão minimizadora de risco por parte dos tomadores de decisão (HARRISON; STEVENS, 1976).

3.4 Métodos de previsão

Os métodos de previsão são ferramentas fundamentais para projetar comportamentos futuros, tendo como propósito a redução do risco na tomada de decisão. Normalmente, as previsões apresentam erros, mas a magnitude dos seus erros depende muito do sistema de previsão utilizado (VERÍSSIMO *et al.*, 2012).

Na Figura 2 são apresentados alguns métodos de previsão de demanda, divididos em dois grupos: os qualitativos, que se baseiam no julgamento de pessoas para determinação de demandas futuras; e os quantitativos, que fazem uso de modelos matemáticos para obtenção de valores previstos.

Segundo Moreira (2008), os métodos quantitativos utilizam modelos matemáticos para realizar as previsões e permitem o controle do erro encontrado. Porém é necessária uma quantidade preliminar de informações para que seja possível a realização da previsão.

Métodos quantitativos são baseados em uma análise dos dados históricos relativos à uma série temporal de variável específica de interesse e possivelmente outras relacionadas à série temporal e examina as relações de causa e efeito da variável com outras variáveis relevantes (MUN, 2010).

Figura 2 - Alguns métodos de previsão



Fonte: Lustosa *et al.* (2008)

3.4.1 Médias Móveis Simples

O método de Médias Móveis Simples é indicado para previsões de curto prazo, onde os componentes tendência e sazonalidade são inexistentes ou possam ser desprezadas (MAKRIDAKIS *et al.*, 1998).

Esse modelo é uma técnica simples de previsão exponencial, onde são considerados os últimos dados históricos e, com estes, é realizada uma média aritmética (ou ponderada) para rever o valor do próximo dado. O número de observações k em cada cálculo da média, ou período, permanece constante e estipulado de maneira a tentar eliminar da melhor forma possível os componentes de tendência e sazonalidade (CHAMBERS *et al.*, 1971; ARCHER, 1980; MAKRIDAKIS *et al.*, 1998).

A equação (1) é a representação matemática do método de Médias Móveis Simples.

$$P_{k+1} = \frac{1}{k} \sum_{i=1}^k Ri \quad (1)$$

onde k é o número de dados passados da série temporal, Ri é o valor real no instante de tempo passado i da série e P_{k+1} é a previsão da série temporal no instante $k + 1$.

3.4.2 Suavização Exponencial Simples

A Suavização Exponencial Simples é um método de fácil implementação e pode ser mais responsivo a novas informações em relação à Média Móvel Simples, considerando a mesma quantidade de dados (SHUMWAY, 2015).

O método de Suavização Exponencial Simples é fundamentado na ponderação de valores analisados da série temporal, com a atribuição de pesos aos dados mais recentes e menores aos mais remotos. Para isso, existe a presença de uma constante de suavização chamada α que varia de 0 a 1.

Na equação (2) é apresentada a representação matemática dada por Makridakis *et al.* (1998) do método de Suavização Exponencial Simples.

$$P_t = \alpha R_{t-1} + (1 - \alpha)P_{t-1} \quad 0 \leq \alpha \leq 1 \quad (2)$$

onde P_t é a previsão de demanda no período t , α é a constante de suavização que recebe um valor que está entre 0 e 1, R_{t-1} é a demanda real no período $t - 1$ e P_{t-1} é a previsão no período $t - 1$.

Utiliza-se um método recursivo para iniciar a série de previsão, que é exemplificado na equação (3).

$$P_1 = R_1 \quad (3)$$

onde P_1 é a previsão de demanda no período 1 e R_1 é a demanda real no período 1.

3.5 Raiz do erro médio quadrático

Veríssimo *et al.* (2012) dizem que normalmente, as previsões apresentam erros, mas a magnitude dos erros depende muito do sistema de previsão utilizado, e afirmam que uma das medidas de erro mais utilizada é a raiz do erro médio quadrático. Segundo Willmott (1981), a raiz do erro médio quadrático é definida pela equação (4).

$$REMQ = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} \quad (4)$$

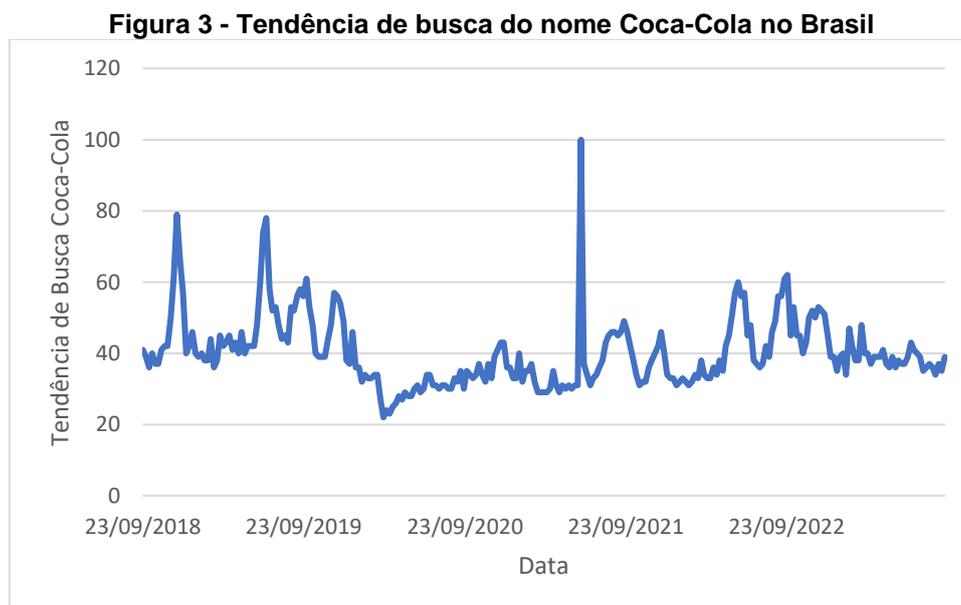
onde e_t é a diferença entre os valores reais e previstos no período t e n é o número de observações da série temporal.

4. Resultados

4.1 Análise dos dados

Com o auxílio do *software* Microsoft Excel e com os dados extraídos do período, foi elaborado um gráfico com as tendências de busca do nome “Coca-Cola” nas plataformas *Google Search*, *Google News* e *Youtube*.

A Figura 3 mostra a série temporal utilizada no trabalho, que representa a tendência de busca do nome Coca-Cola no Brasil.



Fonte: Os autores (2023)

Após isso, foram realizados testes para verificação de estacionariedade da série. Utilizando o *software* R, foi realizado o teste de Dickey-Fuller Aumentado (ADF) e o p-valor é de 0,01755, indicando que a série é estacionária para um valor crítico de 5%. O segundo teste de estacionariedade, o Kwiatkowski-Phillips-Schmidt-Shin (KPSS), foi aplicado utilizando novamente o *software* R, resultando em um p-valor de 0,08987, confirmando a estacionariedade da série ao nível de significância de 5%. É válido ressaltar que não foram excluídos *outliers* ou eventos atípicos.

Também foi realizado o teste de tendência da série por meio do teste de Mann-Kendall, com auxílio do *software* R, encontrou-se o p-valor de 0,7541, indicando que a série não tem tendência, para um valor crítico de 5%.

Como foi observado, ela é uma série estacionária, que não apresenta tendência e sazonalidade, conforme pode-se notar na Figura 2, mas é válido observar o efeito “Coca-Cola” nas buscas. A empresa é uma gigante mundial do comércio de bebidas e seus patrocínios aos grandes eventos tem influência nos dados das buscas, como os Jogos Olímpicos de 2020, realizado em agosto de 2021 e da Copa do Mundo 2022 que foi realizada em novembro/dezembro do mesmo ano, bem como o Natal, onde a campanha de *marketing* dessa empresa é muito forte. Esses períodos ou aqueles que antecedem são os períodos com mais buscas.

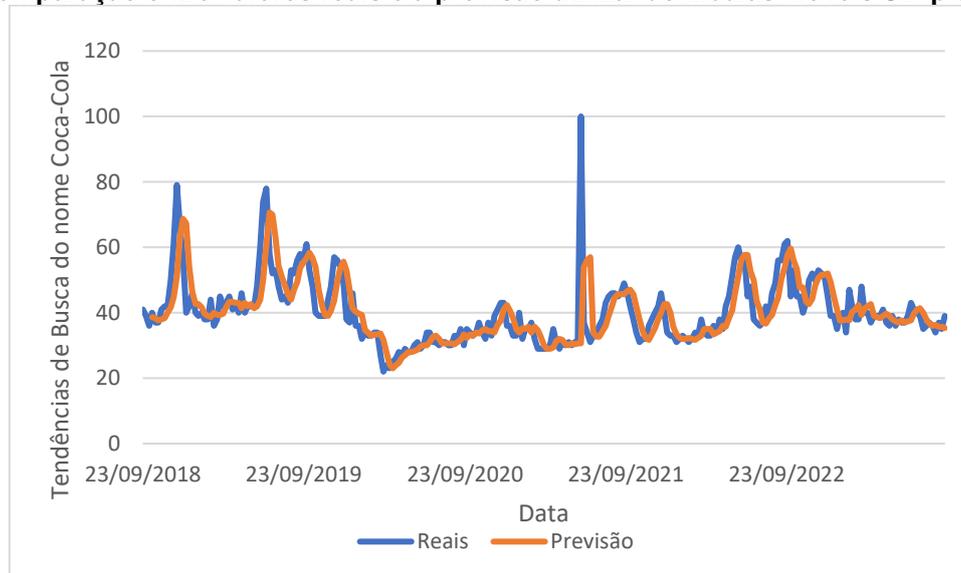
4.2 Aplicação dos Métodos

Médias Móveis Simples

Utilizando a equação (1) no Microsoft Excel, foi possível construir a Figura 4 com as previsões nas tendências de busca, foi utilizada a Médias Móveis Simples com $k = 2$, ou seja, utilizando 2 dados passados da série para realizar a previsão do dado futuro.

Esse modelo foi escolhido, pois se aplica principalmente na ausência de tendência em uma série temporal.

Figura 4 - Comparação entre valores reais e a previsão utilizando Médias Móveis Simples com $k = 2$



Fonte: Os autores (2023)

Foram feitos testes com $k = 1$, $k = 2$ e $k = 3$ e o de $k = 2$ foi o que apresentou o menor erro entre esses três, por isso foi escolhido para a comparação.

Para critério da análise dos erros, foi utilizada a equação (4), utilizando o Microsoft Excel para realizar os cálculos. A Tabela 1 apresenta os erros calculados para os passos supracitados.

Tabela 1 – Erros dos diferentes valores de k para o método de Médias Móveis Simples

k	REMQ
1	7,46891
2	7,32874
3	7,704071

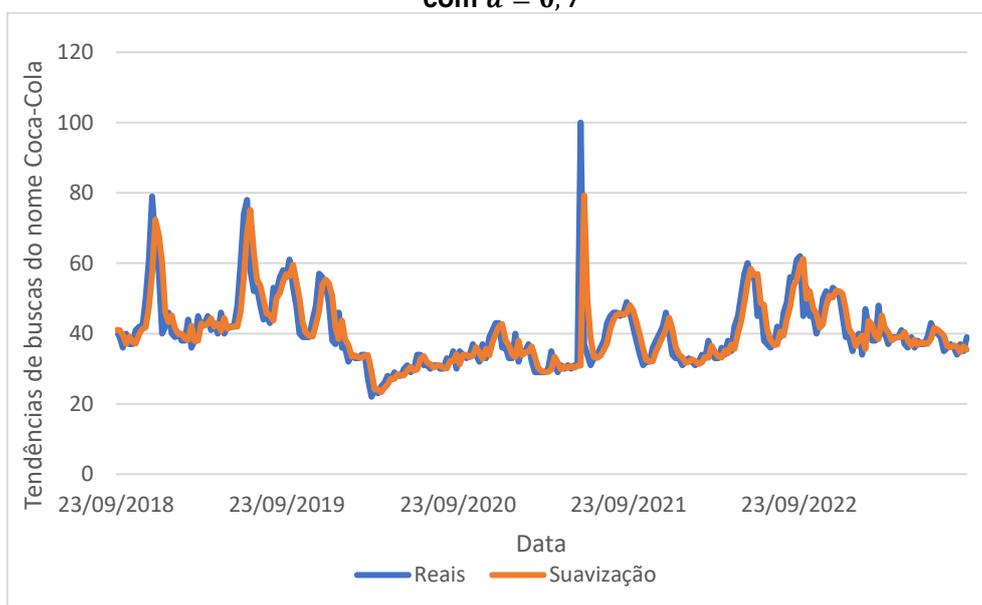
Fonte: Os autores (2023)

Suavização Exponencial Simples

Esse método também foi escolhido por ser adequado para séries estacionárias e que não tenham tendências e nem sazonalidade.

Nessa série, foi utilizada a equação (3) para iniciar as previsões. Utilizando a equação (2) no Microsoft Excel, foi possível construir a Figura 4 com as previsões nas tendências de busca.

Figura 2 - Comparação entre valores reais e previsão utilizando Suavização Exponencial Simples com $\alpha = 0,7$



Fonte: Os autores (2023)

Foram feitos estudos com $\alpha = 0,3$, $\alpha = 0,5$ e $\alpha = 0,7$ e o de $\alpha = 0,7$ foi o que apresentou o menor erro entre esses três, por isso foi escolhido para a comparação.

Para critério da análise dos erros, foi utilizada a equação (4), utilizando o Microsoft Excel para realizar os cálculos. A Tabela 2 apresenta os erros calculados para os passos supracitados.

Tabela 2 – Erros dos diferentes valores de α para o método de Suavização Exponencial Simples

α	REMQ
0,3	8,8904318
0,5	7,17657593
0,7	7,10156911

Fonte: Os autores (2023)

5. Conclusões

A análise comparativa da aplicação dos métodos quantitativos para as previsões de busca na internet no Brasil da marca Coca-Cola mostrou que os dois métodos aplicados tiveram desempenho parecido, principalmente, por estar muito atrelado ao tipo de série que se tem, como dito, a série estudada é uma série estacionária, sem os componentes de tendência e

sazonalidade de séries temporais. Isso facilitou a escolha dos métodos aplicados para realizar as estimativas de previsão e analisar as suas tendências e assertividades.

Os resultados mostram que o método que apresentou o menor erro e o mais indicado a ser utilizado é a Suavização Exponencial Simples com $\alpha = 0,7$, porém os resultados alcançados com Médias Móveis Simples com $k = 2$ ficaram bem próximos ao citado anteriormente.

Esse trabalho pode ajudar a empresa a identificar quais são os períodos em que ela tem menos apelo na internet e tentar prever quais serão seus resultados, caso ela preveja um possível resultado abaixo das expectativas nas buscas, pode realizar ações para que seu desempenho melhore e ela seja mais buscada ou melhor vista pelos consumidores naquele determinado período. Uma grande facilidade dessa pesquisa foi ter a ferramenta *Google Trends* à disposição, o que facilitou a coleta dos dados.

Uma sugestão para trabalhos futuros é a utilização de outros métodos de previsão de demanda, em especial métodos de aprendizado de máquina, que aprendem padrões para projetar futuramente as informações, porém estes demandam grande quantidade de dados no seu treinamento.

Referências

ARCHER, B. Forecasting Demand: Quantitative and Intuitive Techniques. **International Journal of Tourism Management**. USA. v.1, n.1, p. 5-12, 1980.

CHAMBERS, J. C.; MULLICK, S. K.; SMITH, D. D. How to Choose the Right Forecasting Technique. **Harvard Business Review**. USA. v. 49, July-August, p. 45-57, 1971.

GOOGLE. **How Trends Yorks**. Disponível em <<https://newsinitiative.withgoogle.com/pt-br/resources/trainings/google-trends-lesson/>>. Acesso em: 22 set. 2023.

HARRISON, P.J.; STEVENS, C. F. Bayesian Forecasting. **Journal of the Royal Statistical Society**. UK.series B, v. 38, nº 3, p.81-135, 1976.

LUFT, C.P. **Minidicionário Luft**. Ática, 20ª edição, 9ª impressão. São Paulo, 2002.

LUSTOSA, L; MESQUITA, M.A.; QUELHAS, O; OLIVEIRA, R. **Planejamento e Controle da Produção**. Volume 4. ed. Rio de Janeiro: Campus Elsevier, 2008. 376p.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R.J. **Forecasting Methods and Applications**. 3. ed. New York: John Wiley & Sons, 1998.

MOREIRA, D. A. **Administração da Produção e Operações**. 2ª. ed. São Paulo: Cengage Learning, 2008. 624p.

MORETTIN, P.A.; TOLOI, C.M.C. **Análise de Séries Temporais**. São Paulo: Editora Edgard Blucher Ltda., 2004.

MUN, J. **Modeling Risk: Applying Monte Carlo Simulation, Real Options Analysis, Stochastic Forecasting and Portfolio Optimization**. 2ª ed., New York: John Wiley & Sons, 2010.

PINDYCK, R.; RUBENFELD, D.L. **Microeconomia**. 5ª ed., São Paulo: Prentice Hall, 2002.

PEZZOTI, R. **Coca-Cola lidera ranking de marcas mais consumidas do mundo; veja o top 10**. 2022. Disponível em:
<<https://economia.uol.com.br/noticias/redacao/2022/05/26/coca-cola-lidera-ranking-de-marcas-mais-consumidas-do-mundo-veja-o-top-10.htm#:~:text=Pelo%2010%C2%BA%20ano%20seguido%2C%20a,no%20mundo%20to do%20em%202021.>> Acesso em: 22 set. 2023.

RAMSAWAK, R; MOHAN, P; HUTCHINSON, G. Understanding mental health conditions and key coping strategies utilized during major lockdowns in the Carribean based on Google Trends Searches; **Heliyons**, 2023. doi:
<https://doi.org/10.1016/j.heliyon.2023.e19843>.

SPRINGER, S; ZIEGER, M; STRZELECKI A. The rise of infodemiology and infoveillance during COVID-19 crisis. **One Health**, Alemanha, v. 13, 2021.

SHUMWAY, R.H. (2015). **Time Series Analysis and Its Applications With R Examples**, volume 1.

VERÍSSIMO, A.J; ALVES, C.C; HENNING, E; AMARAL, C.E. Métodos estatísticos de suavização exponencial Holt-Winters para previsão de demanda em uma empresa do setor metal mecânico. **Revista Gestão Industrial**. Ponta Grossa, V08, n. 04: p. 154-171, 2012.

WILLMOTT, C. J., **On the validation of models**, *Phys. Geogr.*, 2, 184-194, 1981.