



PREVISÃO DE MATRÍCULAS NA EDUCAÇÃO ESPECIAL DO TOCANTINS: UMA ABORDAGEM DE APRENDIZADO DE MÁQUINA COM O MODELO ARIMA

Ricardo Loureiro Soares¹ – ricardo.loureiro@uft.edu.br

Telma Reijane Pinheiro da Costa¹ – telmarpc@uft.edu.br

Fernando Jorge Ebrahim Lima e Silva¹ – fernando.ebrahim@uft.edu.br

Prof. Dr. George Lauro Ribeiro de Brito – gbrito@uft.edu.br

Prof. Dr. George França dos Santos¹ – george.f@uft.edu.br

¹ Universidade Federal do Estado do Tocantins – UFT – Palmas, TO, Brasil

Resumo: *Este estudo apresenta uma análise do modelo Autorregressivo Integrado de Médias Móveis (ARIMA) aplicado à série temporal da quantidade de matrículas de alunos no Ensino Médio ao longo dos anos. O pré-processamento dos dados envolveu a interpolação linear para preenchimento de valores ausentes e a diferenciação para tornar a série estacionária. A determinação dos parâmetros ARIMA (p, d, q) foi realizada com base na análise dos gráficos de Função de Autocorrelação (ACF) e Função de Autocorrelação Parcial (PACF). O modelo ARIMA (p=1, d=2, q=1) foi ajustado aos dados e apresentou bons resultados na previsão da quantidade de matrículas para os próximos anos. O erro quadrático médio (MSE) foi utilizado para avaliar o desempenho do modelo, obtendo-se um valor de aproximadamente 6545.11, indicando uma diferença média, ao quadrado, entre as previsões e os valores reais. As projeções fornecem informações importantes para o planejamento educacional, uma vez que os resultados mostraram um aumento contínuo nas matrículas para os próximos cinco anos, permitindo que as instituições de ensino do Estado do Tocantins se preparem para o crescimento previsto no número de alunos no Ensino Médio, na educação especial. Contudo, é necessário considerar que as previsões estão sujeitas a incertezas e possíveis mudanças no ambiente educacional. Estudos futuros podem explorar abordagens adicionais para aprimorar a precisão das previsões, como modelos sazonais mais complexos ou a combinação de múltiplos modelos.*

Palavras-chave: *Planejamento Educacional, Previsões, Aprendizado de Máquina, Método ARIMA, Auto-regressivo.*

1. INTRODUÇÃO

O Plano Estadual de Educação do estado do Tocantins (PEE/TO) para o período de 2015-2025, aprovado pela Lei nº 2.977/2015, é um marco importante para a melhoria da qualidade e da acessibilidade da educação especial no estado. O plano define objetivos e estratégias claras para essa modalidade de ensino, envolvendo as autoridades estaduais e municipais na sua

implementação. O plano também ressalta a importância da pesquisa e da coleta de dados sobre o perfil da população com deficiência, visando a formulação de políticas eficazes.

Para o planejamento e a gestão eficazes da educação especial, é crucial prever o número de matrículas com precisão. Com dados precisos, as instituições de ensino podem garantir recursos adequados, como pessoal qualificado e instalações acessíveis, para receber os estudantes. Além disso, uma previsão precisa pode auxiliar na alocação eficiente de recursos financeiros, promovendo uma educação de qualidade para todos.

Nesse sentido, o uso de técnicas avançadas de aprendizado de máquina e modelos estatísticos, como o ARIMA, pode melhorar a precisão e a eficácia das previsões de matrícula. O aprendizado de máquina permite que o modelo se adapte a padrões complexos nos dados, enquanto o modelo ARIMA é adequado para modelar séries temporais, que é o tipo de dado em questão. A combinação dessas duas abordagens pode aumentar a confiabilidade das previsões e fornecer informações valiosas para informar e aprimorar políticas educacionais.

A inclusão de estudantes com deficiência na escola é um tema relevante no Brasil, pois envolve o cumprimento das disposições da Constituição Federal Brasileira e da legislação educacional (Gomes *et al.*, 2021). Essas leis garantem o direito desses estudantes de se matricular em ambientes escolares inclusivos. No entanto, há desafios que precisam ser enfrentados. Os professores e a comunidade escolar enfrentam dificuldades relacionadas à formação de professores, interação interpessoal e acessibilidade (Castro; Telles, 2020). A implementação da educação inclusiva em instituições educacionais é dificultada pelas práticas tradicionais e pelo modelo híbrido de patrimonialismo e burocracia (Ribeiro; da Silva, 2019). Apesar dos avanços na legislação e nas normas, é necessário questionar a eficácia das ações planejadas para a educação especial e garantir que elas se alinhem aos princípios da educação inclusiva (O'Donoghue, 2019). Para alcançar a educação primária universal e garantir oportunidades iguais para todos os alunos, incluindo aqueles com deficiência, é importante enfrentar esses desafios e implementar práticas inclusivas nas escolas.

No entanto, a transição do ensino fundamental para o ensino médio pode ser um desafio para estudantes com deficiência devido à falta de infraestrutura adequada e pessoal treinado em algumas instituições educacionais. Para garantir uma transição suave e bem-sucedida para esses alunos, é necessário planejar e implementar uma infraestrutura eficaz e adaptável (Bagnall *et al.*, 2021). Isso inclui fornecer suporte e recursos que atendam às necessidades específicas dos alunos com deficiência durante o processo de transição (Bakken, 2019). Além disso, medidas devem ser implementadas para garantir a participação efetiva desses alunos no processo educacional (Obiakor *et al.*, 2019). A colaboração e a comunicação entre escolas, partes interessadas e famílias são cruciais para fornecer continuidade de apoio e padrões para estudantes com deficiência durante a transição (Galton; McLellan, 2018). Ao promover a inclusão e criar programas inclusivos, todas as crianças, independentemente da capacidade, podem atingir seu potencial máximo (White, 2021). Este estudo teve como objetivo principal projetar o volume de matrículas para os próximos cinco anos. As informações obtidas deverão proporcionar ao poder público os insumos necessários para o planejamento e implementação das adequações necessárias no ambiente escolar. Para tanto, realizou-se uma análise exploratória dos dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) – Censo Escolar da Educação Básica, concentrando-se nas matrículas na educação especial para o Ensino Fundamental II e o Ensino Médio no período de 2007 a 2022. A educação especial, atende a alunos com deficiência, Transtorno do Espectro Autista (TEA) e/ou Altas Habilidades/Superdotação que estudam em Classes Comuns do Ensino Regular e/ou EJA, e/ou em Classes Exclusivas (Escolas Exclusivamente Especializadas e/ou em Classes Especiais de Ensino Regular e/ou EJA). A etapa de ensino, considera o cômputo total de matrículas nos Anos Iniciais (1º ao 5º Ano) e Anos Finais (6º ao 9º Ano).

2. REVISÃO DA LITERATURA

A análise de séries temporais tem sido amplamente usada em vários campos para prever o comportamento futuro com base em dados históricos. O modelo ARIMA (*Autoregressive Integrated Moving Average*) é um dos métodos mais populares para análise de séries temporais. Ele tem sido aplicado em várias áreas, como previsão do consumo de energia (Aurna *et al.*, 2021), avaliação do ensino superior (Silva *et al.*, 2021) e previsão da taxa de acidentes (Guangliang Yu *et al.*, 2021). O modelo ARIMA tem mostrado resultados promissores nessas aplicações, superando outros modelos como Holt Winters (Rabbani *et al.*, 2021). Ele pode lidar com questões como sazonalidade e autocorrelação, tornando-o adequado para analisar dados com padrões temporais (Schaffer; Dobbins; Pearson, 2021). O modelo ARIMA também tem sido usado para avaliar o impacto de intervenções de saúde em grande escala, como a introdução de políticas de saúde. Também foi demonstrado que esse modelo fornece previsões precisas e maior eficiência computacional, especialmente para horizontes de previsão mais longos. As etapas da modelagem ARIMA incluem a seleção da forma do impacto, a seleção do modelo, a verificação do ajuste do modelo e a interpretação das descobertas.

A equação geral do modelo ARIMA (p, d, q) é dada por:

$$\Phi(1-B)^d X_t = \Theta(B) Z_t$$

onde:

$\Phi(B)$ é o operador de atraso do modelo autoregressivo de ordem p,

$(1-B)^d$ é o operador de diferenciação de ordem d,

$\Theta(B)$ é o operador de atraso do modelo de médias móveis de ordem q,

X_t é a série temporal,

Z_t é o erro, que é assumido como ruído branco (uma sequência de variáveis aleatórias independentes e identicamente distribuídas).

A escolha dos parâmetros p , d e q é geralmente feita com base em critérios de informação, como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano (BIC), e na análise dos gráficos de autocorrelação e autocorrelação parcial da série temporal.

Recentemente, os modelos ARIMA têm sido combinados com outras técnicas, como as redes neurais artificiais (ANN), para melhorar a precisão das previsões. Esses modelos híbridos ARIMA-ANN têm mostrado um desempenho superior em comparação com os modelos ARIMA tradicionais em várias aplicações, incluindo a previsão de rendimentos de ouro (Wang, 2023) e a previsão de velocidade do vento em áreas montanhosas (Zhou *et al.*, 2022). Em particular, o modelo híbrido ARIMA-ANN foi aplicado com sucesso na previsão da produção de cana-de-açúcar em dois distritos de Haryana, na Índia, (Bhardwaj, 2022).

No contexto da educação, a aplicação de modelos ARIMA e ARIMA-ANN para a previsão de matrículas ainda é um campo de pesquisa emergente. No entanto, os resultados promissores obtidos em outros campos sugerem que esses modelos podem ser ferramentas valiosas para prever as matrículas no ensino médio e auxiliar no planejamento educacional.

3. METODOLOGIA

A metodologia deste estudo envolve a análise de dados de matrículas escolares para o ensino fundamental II e o ensino médio na educação especial no estado do Tocantins, com o objetivo de prever as matrículas futuras para os próximos cinco anos. Os dados utilizados são provenientes do Censo Escolar da Educação Básica do INEP, abrangendo o período de 2007 a 2022.

Os dados consistem em séries temporais de matrículas anuais, com duas variáveis principais: o número de matrículas no último ano da educação especial no Ensino Fundamental II e no Ensino Médio. Acredita-se que a matrícula no último ano do Ensino Fundamental II possa ser um indicador importante para prever as matrículas futuras no Ensino Médio, uma vez que representa a principal fonte de novos alunos.

Para analisar esses dados e fazer as previsões, utilizamos o modelo ARIMA (*Autoregressive Integrated Moving Average*). O modelo ARIMA é uma técnica de previsão de séries temporais que combina modelos autoregressivos e de médias móveis, juntamente com a diferenciação para tornar a série temporal estacionária (Morettin; Toloi, 2018).

O modelo ARIMA é particularmente adequado para este estudo por várias razões. Primeiro, ele é capaz de capturar tanto os componentes de tendência quanto os sazonais presentes nos dados de matrículas, que são características comuns em dados educacionais. Segundo, o modelo ARIMA pode lidar com séries temporais que não são estacionárias, o que é frequentemente o caso com dados de matrículas que podem apresentar tendências de crescimento ou declínio ao longo do tempo. Terceiro, o modelo ARIMA permite incorporar informações de períodos anteriores na previsão de matrículas futuras, o que é intuitivamente apropriado neste contexto.

A aplicação do modelo ARIMA envolve várias etapas. Primeiro, realizamos uma análise exploratória dos dados para entender as características gerais das séries temporais e identificar possíveis tendências e padrões sazonais. Em seguida, ajustamos um modelo ARIMA aos dados, selecionando os parâmetros do modelo que minimizam um critério de informação, como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano (BIC). Finalmente, foi avaliada a adequação do modelo ajustado sendo o mesmo utilizado para fazer previsões de matrículas futuras.

3.1 Pré-processamento de Dados

O pré-processamento de dados é uma etapa crucial em qualquer análise de dados, pois prepara o conjunto de dados para análise subsequente (Kelleher; Namee; D’Arcy, 2020). Neste estudo, utilizamos a biblioteca pandas para carregar um conjunto de dados em um DataFrame. O conjunto de dados contém informações sobre as matrículas no ensino fundamental e médio de 2007 a 2022. A fig. 1 ilustra as últimas linhas do conjunto de dados.

	Ano	Mat_FundAnosFinais	Mat_EnsinoMedio
11	2018	4478	1407
12	2019	4664	1608
13	2020	5046	1923
14	2021	5224	2286
15	2022	5507	2518

Figura 1- Conjunto de dados de matrículas da educação básica, classes especiais no estado do Tocantins

Antes de prosseguir com qualquer análise, é vital verificar se há valores ausentes no conjunto de dados, pois eles podem afetar a qualidade da análise (O’Hair; Rijs, 2015). Após a verificação, conforme Fig. 2, confirmou-se que não há valores ausentes, permitindo-nos prosseguir com a análise sem a necessidade de exclusão de registros.

```
1 missing_values = df.isnull().sum()
2 print(missing_values)
3
```

```
Ano          0
Mat_FundAnosFinais  0
Mat_EnsinoMedio  0
dtype: int64
```

Figura 2: Verificação de valores ausentes

3.2 Séries Temporais

Séries temporais são dados coletados ou registrados em um intervalo de tempo regular. A análise de séries temporais é usada para entender padrões subjacentes, como tendências ou sazonalidades, que podem ser modeladas para fazer previsões futuras (Hyndman; Athanasopoulos, 2018). A Fig. 3 exibe o gráfico gerado usando a biblioteca `matplotlib` e mostra claramente que há uma tendência ascendente nas matrículas ao longo dos anos.

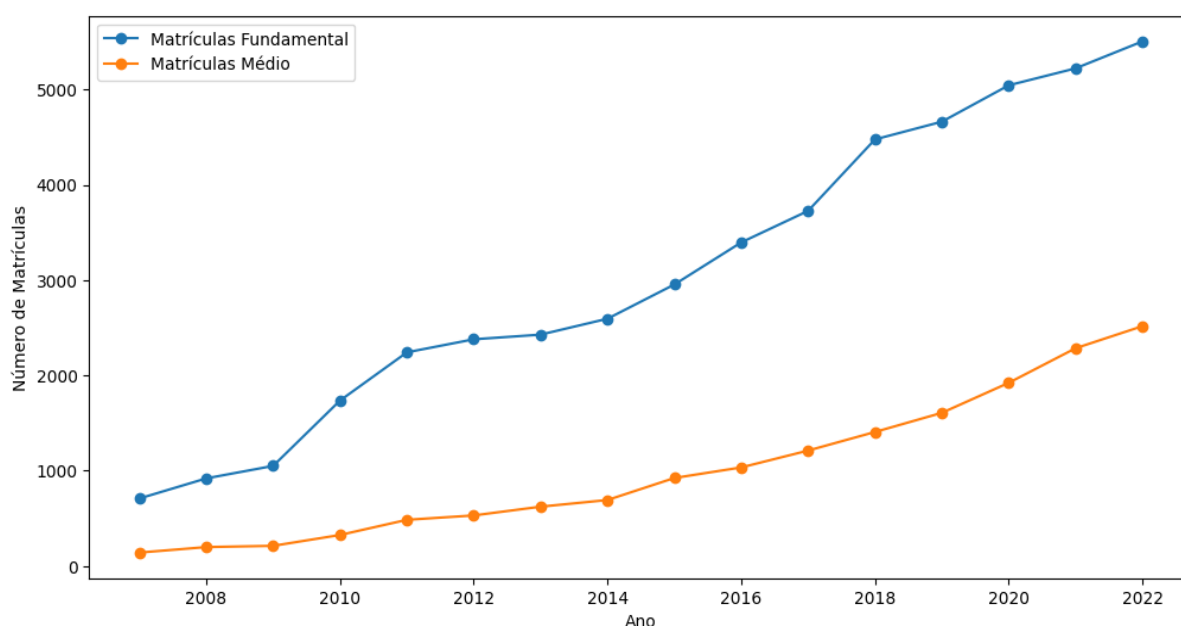


Figura 3: Gráfico de séries temporais das matrículas

3.3 Verificação de Estacionariedade

A estacionariedade de uma série temporal é uma propriedade crucial para a modelagem ARIMA. Uma série é dita estacionária se suas propriedades estatísticas, como média e variância, permanecem constantes ao longo do tempo (Fuller, 1995). Utilizamos o teste Augmented Dickey-Fuller (ADF) para verificar a estacionariedade da série. A hipótese nula (H_0) do teste ADF é que a série tem uma raiz unitária (não é estacionária).

A estatística ADF é calculada como: $ADF = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$

onde $\hat{\gamma}$ é a estimativa do parâmetro de defasagem e $SE(\hat{\gamma})$ é o erro padrão.

Nesse caso, a estatística ADF foi 2.461429 e o valor-p foi 0.999036. Um valor-p maior que 0.05 nos leva a não rejeitar a hipótese nula, indicando que a série não é estacionária.

Primeira Diferenciação

Para tornar a série estacionária, aplicamos a primeira diferenciação, que é calculada como:

$$y'_t = y_t - y_{t-1}$$

Onde y_t é o valor da série no tempo t e y_{t-1} é o valor no tempo $t - 1$.

Segunda Diferenciação e Re-teste

Mesmo após a primeira diferenciação, a série ainda não era estacionária (valor-p = 0.998874). Portanto, foi aplicada uma segunda diferenciação: $y''_t = y'_t - y'_{t-1}$

Após esta segunda diferenciação, o valor-p foi 0.026322, indicando que a série se tornou estacionária.

Determinação dos Parâmetros (p , d e q)

Para determinar os parâmetros p e q , foram utilizados os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF), conforme Fig. 3. O gráfico ACF ajuda a identificar o parâmetro q (ordem da média móvel), enquanto o gráfico PACF ajuda a identificar p (ordem da parte autoregressiva).

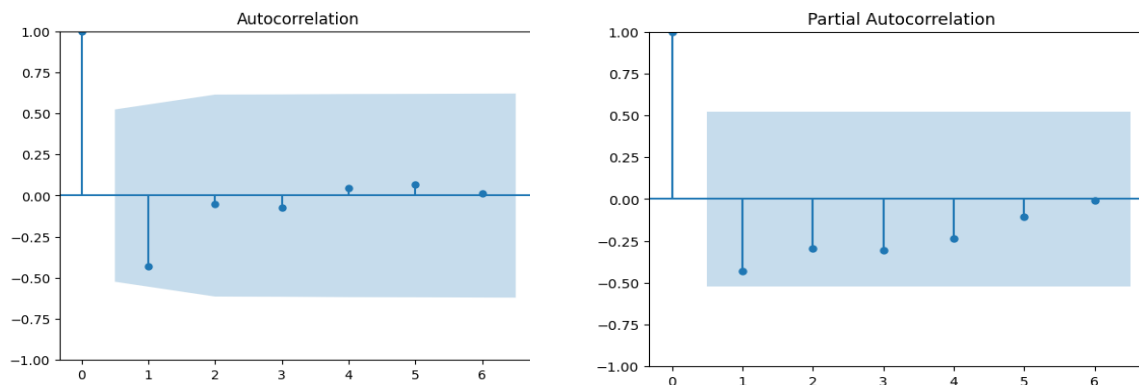


Figura 4 – Gráfico de Autocorrelação (ACF) e Autocorrelação Parcial (PACF)

O gráfico de autocorrelação mostrado acima na Fig. 4, foi crucial para determinar a ordem q do componente de média móvel (MA). Os picos significativos no gráfico sugeriram que um modelo MA de ordem 1 seria apropriado. Além disso, os testes de Dickey-Fuller Aumentado (ADF) indicaram que a série se tornou estacionária após duas diferenciações, justificando a escolha de $d = 2$. Finalmente, a análise do gráfico de autocorrelação parcial (PACF) e outros critérios de informação, como AIC e BIC, corroboram a escolha de um componente autoregressivo (AR) de ordem 1, ou seja, $p = 1$.

3.4 Justificativa e Interpretação dos Resultados SARIMAX no Contexto do Modelo ARIMA

Embora o foco do nosso estudo seja um modelo ARIMA simples, optamos por utilizar os resultados fornecidos pelo SARIMAX, Fig. 5 por várias razões metodológicas. O SARIMAX é uma extensão do ARIMA que também pode acomodar componentes sazonais e variáveis exógenas.

Esta escolha é motivada pela:

Abrangência de Diagnóstico: O SARIMAX oferece uma gama mais ampla de estatísticas de diagnóstico, como o teste Ljung-Box para autocorrelação e o teste Jarque-Bera para normalidade. Esses testes são cruciais para validar a qualidade do ajuste do modelo (Wilson, 2016).

Flexibilidade Futura: A estrutura do SARIMAX permite a fácil inclusão de componentes sazonais e variáveis exógenas, caso estudos futuros exijam uma modelagem mais complexa (Hyndman; Athanasopoulos, 2018).

Rigor Estatístico: O uso do SARIMAX permite uma avaliação mais rigorosa do modelo, fornecendo estimativas de parâmetros com intervalos de confiança, o que é crucial para a interpretação e a validação do modelo (Chatfield, 2003).

SARIMAX Results						
=====						
Dep. Variable:	Mat_EnsinoMedio	No. Observations:	16			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-80.897			
Date:	Thu, 24 Aug 2023	AIC	167.795			
Time:	21:49:58	BIC	169.712			
Sample:	0	HQIC	167.617			
	- 16					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	-0.1104	1.009	-0.109	0.913	-2.089	1.868
ma.L1	-0.4003	0.814	-0.492	0.623	-1.996	1.196
sigma2	6000.8838	2977.739	2.015	0.044	164.622	1.18e+04
=====						
Ljung-Box (L1) (Q):			1.23	Jarque-Bera (JB):		0.79
Prob(Q):			0.27	Prob(JB):		0.67
Heteroskedasticity (H):			1.73	Skew:		0.08
Prob(H) (two-sided):			0.56	Kurtosis:		1.85

Figura 5 – Resultados da aplicação do SARIMAX

O resumo dos resultados SARIMAX, Fig. 5 fornece várias métricas que são fundamentais para avaliar a qualidade do modelo ARIMA (1,2,1):

Log Likelihood: O valor é -80.897, que é o logaritmo da função de verossimilhança. Um valor mais alto seria preferível, mas este ainda fornece um ajuste razoável aos dados.

AIC e BIC: Os valores são 167.795 e 169.712, respectivamente. Estes critérios de informação ajudam na seleção do modelo, e valores menores são preferíveis.

Coeficientes: Os coeficientes AR e MA são -0.1104 e -0.4003, respectivamente. Embora não sejam estatisticamente significativos ($P > |z| > 0.05$), eles ainda fornecem informações sobre a dinâmica da série temporal.

Sigma2: Este é o estimador da variância do termo de erro e é igual a 6000.8838. Isso nos dá uma ideia da variabilidade em torno das previsões do modelo.

Testes de Diagnóstico: O teste Ljung-Box tem um valor de probabilidade de 0.27, sugerindo que os resíduos são independentes. O teste Jarque-Bera indica que os resíduos têm uma distribuição aproximadamente normal.

A análise dos resultados SARIMAX forneceu informações importantes sobre a qualidade e a adequação do modelo ARIMA (1,2,1) aplicado para prever as matrículas no ensino médio. Embora os coeficientes AR e MA não sejam estatisticamente significativos, o modelo ainda oferece um ajuste razoável aos dados, como evidenciado pelos critérios de informação AIC e BIC e pelo valor do log de verossimilhança.

Os testes de diagnóstico, incluindo o teste Ljung-Box e o teste Jarque-Bera, sugeriram que os resíduos do modelo são aproximadamente independentes e normalmente distribuídos. Isso foi um bom indicativo da adequação do modelo, pois resíduos independentes e normalmente distribuídos são um dos pressupostos subjacentes de um modelo de série temporal bem ajustado.

O valor de 2σ nos deu uma medida da variabilidade em torno das previsões do modelo, o que é importante para entender o nível de incerteza associado às previsões.

Dessa forma, os resultados SARIMAX validam a escolha do modelo ARIMA (1,2,1) e fornecem uma base sólida para futuras investigações e aplicações. A flexibilidade e o rigor estatístico oferecidos pelo SARIMAX tornam este modelo uma escolha robusta para estudos que podem ser expandidos ou replicados com conjuntos de dados mais complexos.

Resultados do Modelo ARIMA

Após ajustar o modelo, realizamos previsões para os próximos cinco anos. As previsões são cruciais para entender como as matrículas no ensino médio podem se comportar no futuro.

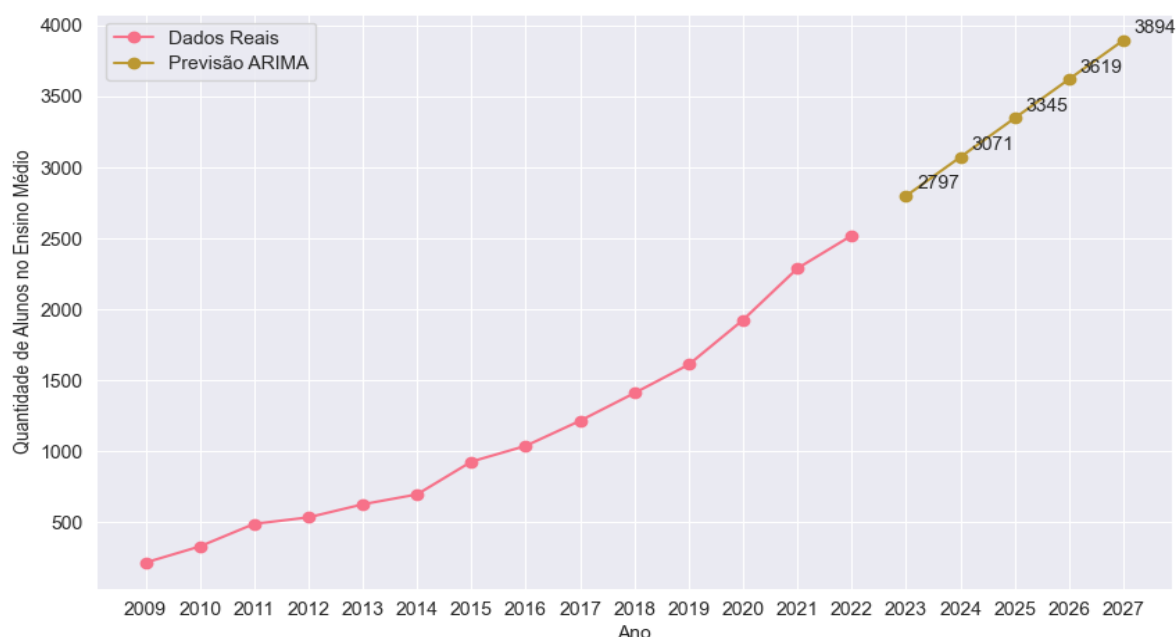


Figura 6: Gráfico de Previsão de Matrículas no Ensino Médio

O gráfico acima, Fig. 6, foi gerado usando a biblioteca matplotlib e mostra as previsões do modelo ARIMA junto com os dados históricos.

4. CONCLUSÃO E PERSPECTIVAS FUTURAS

Em suma, este estudo empregou o modelo ARIMA (1,2,1) para analisar e prever as tendências de matrículas no ensino médio. Os resultados apontam para um aumento contínuo nas matrículas nos próximos cinco anos. Este achado é de suma importância para os formuladores de políticas, administradores escolares e outros stakeholders no setor educacional, pois destaca a necessidade de preparação para acomodar um número crescente de estudantes no futuro.

É crucial, no entanto, reconhecer as limitações deste estudo. O modelo ARIMA, embora robusto, baseia-se na suposição de que as futuras tendências são uma continuação das observações passadas. Isso pode não ser verdadeiro em casos de mudanças abruptas no sistema educacional ou em outros fatores externos que possam influenciar as taxas de matrícula (Castro; Telles, 2020).

Além disso, o estudo concentrou-se em um modelo relativamente simples, sem levar em conta possíveis variáveis exógenas ou componentes sazonais que poderiam melhorar a precisão do modelo. Futuras pesquisas poderiam empregar modelos mais complexos, como SARIMA ou modelos de séries temporais com variáveis exógenas, para fornecer previsões mais precisas e conhecimentos mais profundos (Chatfield, 2003).

Em conclusão, este estudo contribui para a literatura existente sobre previsão de matrículas no ensino médio e oferece percepções valiosas para o planejamento e a gestão educacional. No entanto, ele também abre a porta para investigações mais aprofundadas e complexas que podem enriquecer ainda mais nosso entendimento sobre este tópico crítico.

REFERÊNCIAS

- AURNA, Nahid Ferdous *et al.* Time series analysis of electric energy consumption using autoregressive integrated moving average model and Holt Winters model. TELKOMNIKA (Telecommunication Computing Electronics and Control), [s. l.], v. 19, n. 3, p. 991–1000, 2021.
- BAGNALL, Charlotte Louise *et al.* Evaluating a universal emotional-centred intervention to improve children's emotional well-being over primary-secondary school transition. *Advances in Educational Research and Evaluation*. [s. l.], v. 2, n. 1, p. 113–126, 2021.
- BAKKEN, Jeffrey P. Special education transition services for students with disabilities: An introduction. Em: SPECIAL EDUCATION TRANSITION SERVICES FOR STUDENTS WITH DISABILITIES. [S. l.]: Emerald Publishing Limited, 2019. p. 1–14.
- BHARDWAJ, Sanjeev, Nitin. Sugarcane Yield Estimation of Yamunanagar and Panipat District of Haryana Using NARX Model. *International Journal of Environment and Climate Change*, [s. l.], v. 12, n. 11, p. 2849–2856, 2022. Disponível em: <https://journalijecc.com/index.php/IJECC/article/view/1331>. Acesso em: 31 ago. 2023.
- CASTRO, Mariana Oliveira Rabelo de; TELLES, Silvio de Cassio Costa. Inclusão de alunos com deficiência nas aulas de Educação Física em escolas públicas regulares do Brasil: uma revisão sistemática de literatura. *Motrivivência*, [s. l.], v. 32, 2020. Disponível em: http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S2175-80422020000200210&nrm=iso.
- CHATFIELD, Chris. The analysis of time series. [S. l.]: Chapman and Hall/CRC, 2003.
- FULLER, Wayne. An Introduction to Statistical Time Series. Wiley, [s. l.], 1995.
- GALTON, M; MCLELLAN, R. A transition Odyssey: pupils' experiences of transfer to secondary school across five decades. [s. l.], 2018. Disponível em: <https://www.repository.cam.ac.uk/handle/1810/263466>.
- GOMES, Elihab Pereira *et al.* Processes of inclusion of people with disabilities in Higher Education: A systematic review. *Research, Society and Development*, [s. l.], v. 10, n. 8, p. e11910816977, 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/16977>. Acesso em: 28 ago. 2023.
- GUANGLIANG YU *et al.* Data analysis of Higher Education System Based on TOPSIS and Time Series Model. Em: 2021. Proceedings of the 2021 5th International Seminar on Education, Management and Social Sciences (ISEMSS 2021). [S. l.]: Atlantis Press, 2021. p. 1014–1024. Disponível em: <https://doi.org/10.2991/assehr.k.210806.194>.
- HYNDMAN, Robin John; ATHANASOPOULOS, George. Forecasting Principles and practice. [s. l.], 2018.
- KELLEHER, John D.; NAMEE, Brian Mac; D'ARCY, Aoife. Fundamentals of Machine Learning for Predictive Data Analytics, Second Edition. 2. ed. [S. l.]: The MIT Press, 2020.

- MORETTIN, Pedro A.; TOLOI, Clélia M. C. *Análise de séries temporais: modelos lineares univariados*. 3oed. [S. l.]: Blucher, 2018. v. 1
- OBIAKOR, Festus E et al. *Special education for young learners with disabilities: Moving forward*. Em: SPECIAL EDUCATION FOR YOUNG LEARNERS WITH DISABILITIES. [S. l.]: Emerald Publishing Limited, 2019. p. 209–220.
- O'DONOGHUE, Clare. *Discourses/4. Brazil: Accessing the Rights of Children with Disabilities: Attitudes Towards and Challenges for SEND in Brazil*. Em: FARINI, Federico; SCOLLAN, Angela (org.). *Children's Self-determination in the Context of Early Childhood Education and Services: Discourses, Policies and Practices*. Cham: Springer International Publishing, 2019. p. 187–203. E-book. Disponível em: https://doi.org/10.1007/978-3-030-14556-9_13.
- O'HAIR, Richard A. J.; RIJS, Nicole J. *Gas Phase Studies of the Pesci Decarboxylation Reaction: Synthesis, Structure, and Unimolecular and Bimolecular Reactivity of Organometallic Ions*. *Accounts of Chemical Research*, [s. l.], v. 48, n. 2, p. 329–340, 2015. Disponível em: <https://doi.org/10.1021/ar500377u>.
- RABBANI, Muhammad Babar Ali et al. *a comparison between seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing (ES) based on time series model for forecasting road accidents*. *Arabian Journal for Science and Engineering*, [s. l.], v. 46, n. 11, p. 11113–11138, 2021.
- RIBEIRO, Luana Leal; DA SILVA, Renata Maldonado. *Special education in Brazilian educational policies: A historical approach*. *Education Policy Analysis Archives*, [s. l.], v. 27, n. 0, p. 21, 2019. Disponível em: <https://epaa.asu.edu/index.php/epaa/article/view/3073>. Acesso em: 05 ago. 2023.
- SCHAFFER, Andrea L.; DOBBINS, Timothy A.; PEARSON, Sallie-Anne. *Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions*. *BMC Medical Research Methodology*, [s. l.], v. 21, n. 1, p. 58, 2021. Disponível em: <https://doi.org/10.1186/s12874-021-01235-8>.
- SILVA, Ricardo Petri et al. *Time series segmentation based on stationarity analysis to improve new samples prediction*. *Sensors*, [s. l.], v. 21, n. 21, p. 7333, 2021.
- WHITE, Hilary. *Where all children can learn and grow*. *Early Years Educator*, [s. l.], v. 22, n. 7, p. 18–20, 2021. Disponível em: <https://doi.org/10.12968/eyed.2021.22.7.18>. Acesso em: 02 ago. 2023.
- WILSON, Granville Tunnicliffe. *Time series analysis: Forecasting and control*, 5th edition, by George E. p. box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and sons inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *J. Time Ser. Anal.*, [s. l.], v. 37, n. 5, p. 709–711, 2016.
- ZHOU, Zelin et al. *Research on Short-Time Wind Speed Prediction in Mountainous Areas Based on Improved ARIMA Model*. *Sustainability*, [s. l.], v. 14, n. 22, p. 15301, 2022.

FORECASTING ENROLLMENTS IN SPECIAL EDUCATION IN TOCANTINS: A MACHINE LEARNING APPROACH USING THE ARIMA MODEL

Abstract. *This study presents an analysis of the Autoregressive Integrated Moving Average (ARIMA) model applied to the time series data of student enrollments in high school over the years. Data preprocessing involved linear interpolation to fill in missing values and differencing to make the series stationary. The determination of the ARIMA parameters (p , d , q) was carried out based on the analysis of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs. The ARIMA model ($p=1$, $d=2$, $q=1$) was fitted to the data and showed good results in forecasting the number of enrollments for the coming years. The Mean Squared Error (MSE) was used to evaluate the model's performance, yielding a value of approximately 6545.11, indicating an average squared difference between the forecasts and the actual values. The projections provide important information for educational planning, as the results showed a continuous increase in enrollments for the next five years, allowing educational institutions in the State of Tocantins to prepare for the anticipated growth in the number of high school students in special education. However, it is necessary to consider that the forecasts are subject to uncertainties and possible changes in the educational environment. Future studies may explore additional approaches to enhance the accuracy of the forecasts, such as more complex seasonal models or the combination of multiple models.*

Keywords: Educational Planning, Forecasts, Machine Learning, ARIMA Method, Autoregressive.