

Revisão do Pré-projeto

Disciplina: Trabalho de Conclusão de Curso I – BCC

Caro, orientando,

segue abaixo o Termo de Compromisso, as DUAS revisões do seu pré-projeto contendo a avaliação do professor “avaliador” e professor “TCC1”, junto com as avaliações da defesa na banca de qualificação. É muito importante que revise com cuidado e discuta possíveis dúvidas decorrente das revisões com o seu professor orientador, e com o professor de TCC1. Sempre procure fazer todos os ajustes solicitados, até mesmo os menores detalhes, pois todos são importantes e irão refletir na sua nota nesta disciplina. Lembre de abrir localmente em um visualizador PDF para poder ver as anotações que foram feitas. E, aparecendo uma anotação feita por mim (prof. De TCC1) que inicie com “TF-...” (ex. “TF-ALÍNEA”) se refere a ajustes de formatação indicando que deve usar o estilo do Word correto do modelo do projeto.

Mas, caso o professor orientador julgue que algumas anotações das revisões não devam ser feitas, ou mesmo que sejam feitas de forma diferente a solicitada pelo revisor, anexe ao final do seu projeto a ficha “Projeto: Observações – Professor Orientador” disponível no material da disciplina, e justifique o motivo.

Lembrem que agora o limite de páginas do projeto é no máximo 12 (doze) páginas. E que a seção de “Revisão Bibliográfica” deve ser complementada (ampliada) usando subseções para cada assunto apresentado.

Atenciosamente,

TERMO DE COMPROMISSO - BCC

I – IDENTIFICAÇÃO DO ALUNO	
Nome:	Rossana Ariadna Schumann Dullius
Telefone:	(47) 997419246
II – IDENTIFICAÇÃO DO TRABALHO	
Título:	PROJEÇÕES DE MATRÍCULAS PARA O ENSINO MÉDIO UTILIZANDO SÉRIES TEMPORAIS
Orientador:	Aurélio Faustino Hoppe
Coorientador (se houver):	
III – COMPROMISSO DE REALIZAÇÃO DO TCC	
Eu (aluno), <u>Rossana Ariadna Schumann Dullius</u> comprometo-me a realizar o trabalho proposto no semestre <u>2023/02</u> , de acordo com as normas e os prazos determinados pela FURB, conforme previsto na resolução nº.20/2016.	
IV – COMPROMISSO DE ORIENTAÇÃO	
Eu (orientador), <u>Aurélio Faustino Hoppe</u> comprometo-me a orientar o trabalho proposto no semestre <u>2023/02</u> , de acordo com as normas e os prazos determinados pela FURB, conforme previsto na resolução nº.20/2016.	

CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC		
(X) PRÉ-PROJETO	() PROJETO	ANO/SEMESTRE: 2023/1

PREVISÃO DO NÚMERO DE MATRÍCULAS PARA O ENSINO MÉDIO DE BLUMENAU UTILIZANDO SÉRIES TEMPORAIS

Rossana Ariadna Schumann Dullius

Prof. Aurélio Faustino Hoppe – Orientador

1 INTRODUÇÃO

O contexto educacional do brasileiro ainda é desafiador, e a demanda por infraestrutura educacional é uma das principais questões a serem enfrentadas. Segundo Souza (2019), a precariedade da infraestrutura das escolas é um obstáculo para a qualidade da educação no país. A infraestrutura precária dificulta a inclusão de tecnologia e de atividades extracurriculares nas escolas, também a falta de investimento em reformas e manutenção de prédios escolares resulta em problemas estruturais, como salas superlotadas, falta de acessibilidade, pouca ventilação, dentre outros.

Segundo Machado e Falsarella (2020), a demanda escolar é uma preocupação constante de gestão pública no Brasil pois ela é responsável por planejar e gerenciar a oferta de vagas escolares, considerando as particularidades locais e as demandas dos estudantes. Ainda de acordo com os autores, neste caso, é importante que a gestão pública acompanhe a evolução da demanda escolar, para tanto é fundamental o investimento em políticas públicas que ampliem a oferta de vagas, como a construção de novas escolas e a ampliação das existentes, além das formas de incentivar a matrícula e permanência dos estudantes na escola.

No que se refere aos números totais de matrículas, segundo Instituto Nacional de Estudos e Pesquisas Educacionais (2022), o ensino básico como um todo, incluindo a educação de jovens e adultos, somou 35.760.081 matrículas em 2022, o que representa uma redução de 2,07% em relação aos dados preliminares de 2021. Esses dados são importantes não só para a gestão educacional, mas também para a distribuição de recursos do Fundo de Manutenção e Desenvolvimento da Educação Básica (Fundeb) e para a execução de programas na área da educação.

De acordo com Bodewe (2022), para realizar a projeção de matrículas com precisão, é necessário considerar diversas variáveis que influenciam o número de alunos matriculados, como número de nascimentos, migração de alunos entre escolas e evasão escolar. Além disso, é importante utilizar técnicas de análise de dados adequadas e contar com profissionais capacitados para realizar as análises. Dentre as técnicas de análise de séries temporais, destacam-se a análise de tendências e a análise de sazonalidade.

Neste contexto, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) disponibiliza diversas ferramentas para análise de dados educacionais, incluindo o Consulta Matrícula, que apresenta o quantitativo de matrículas declaradas ao Censo Escolar anualmente e exibe um comparativo com o ano anterior (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS, 2020). Essa ferramenta pode ser útil para a realização de análises de séries temporais e para a projeção de matrículas no ensino médio. No entanto, é importante destacar que a predição de matrículas é uma atividade complexa que exige alto grau de precisão e confiabilidade. Por isso, é recomendável utilizar outras fontes de dados e técnicas de análise de dados para complementar a análise (NEVES; RIGOTTI, 2018).

De acordo com (GARTNER, 2023), a predição é um processo analítico que utiliza modelos estatísticos e algoritmos de aprendizado de máquina para prever eventos futuros com base em dados históricos. Essa técnica é amplamente utilizada em diversos campos, incluindo negócios, finanças, medicina, engenharia, entre outros. Além disso, a predição também pode auxiliar no processo de projeções, permitindo que as organizações prevejam demanda futura e tomem decisões estratégicas.

A partir deste contexto, surge a seguinte pergunta de pesquisa: a utilização de modelos preditivos em conjunto com séries temporais pode auxiliar nas projeções da demanda municipal de escolas e professores no ensino médio?

1.1 OBJETIVOS

O objetivo deste trabalho é analisar como a utilização de modelos preditivos em conjunto com séries temporais pode contribuir para projeções precisas da demanda municipal de escolas e professores no ensino médio.

Os objetivos específicos são:

- coletar e analisar os dados históricos de matrículas no ensino médio em Blumenau;
- identificar as variáveis relevantes para a predição da demanda de matrículas, como o número de alunos ingressantes no ensino fundamental e o índice de natalidade;

Comentado [JM1]: Excluir: "do"

Comentado [JM2]: Excluir: ", "

Comentado [JM3]: Trocar: "também"

Por: "assim como"

Comentado [JM4]: Trocar: "demanda escolar, para tanto"

Por: "demanda escolar. Para tanto,"

Comentado [JM5]: Trocar: "como"

Por: "com"

Comentado [JM6]: Trocar: "Dentre as técnicas de análise de séries temporais, destacam-se"

Por: "Dentre as técnicas de análise que podem ser utilizadas, tem-se as séries temporais, destacando-se"

Comentado [JM7]: Trocar: "De acordo com (GARTNER, 2023),"

Por: "De acordo com Gartner (2023),"

Comentado [JM8]: Trocar: "demanda municipal de escolas e professores do ensino médio?"

Por: "demanda municipal de escolas e professores do ensino médio em Blumenau?"

Dúvida: você vai verificar demanda por estudantes e professores?

Comentado [JM9]: Quais?

Comentado [JM10]: Trocar: "professores no ensino médio."

Por: "professores no ensino médio em Blumenau."

- c) aplicar técnicas de análise de séries temporais para estimar o número de matrículas no ensino médio considerando cenários de curto, médio e longo prazo;
- d) avaliar a precisão e confiabilidade do modelo desenvolvido.

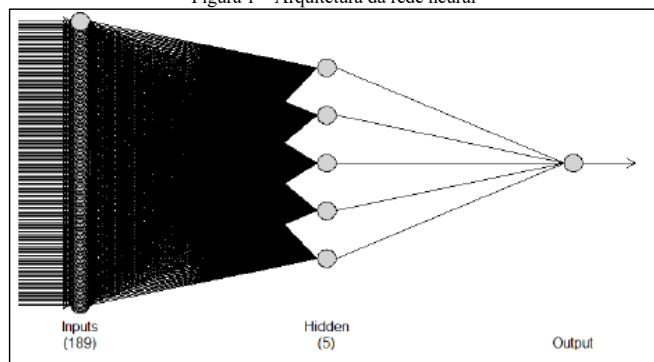
2 TRABALHOS CORRELATOS

Nesta seção serão apresentados três trabalhos que possuem semelhanças com o trabalho proposto. Na subseção 2.1 encontra-se o trabalho de Piva (2021), ao qual utiliza rede neural a fim de criar um modelo de previsão para suprir a demanda escolar. Na subseção 2.2 descreve-se a aplicação do modelo de Monte Carlo para previsão de matrículas para gerar estratégias a uma Instituição de Ensino Superior (IES) (PANDOLFI *et al.*, 2014). Por fim, na subseção 2.3 detalha-se o trabalho de Cruz *et al.* (2020) que apresentam o uso do algoritmo ARIMA para a realização de previsão de matrículas em uma IES.

2.1 MODELO DE PREVISÃO DE VAGAS DE EDUCAÇÃO INFANTIL DE UM MUNICÍPIO BRASILEIRO

Piva (2021) desenvolveu um modelo utilizando a arquitetura de rede neural multicamadas visando prever a demanda da educação infantil em um município brasileiro de Camboriú. Para isso, o autor coletou informações dos registros para a fila de espera do sistema municipal. Posteriormente, formulou a série temporal com os inscritos da fila de espera. Para implementação, Piva (2021) utilizou a biblioteca *nnfor* da linguagem R. Os parâmetros utilizados foram definidos utilizando o método proposto por Crone e Kourentzes (2010), onde estima-se os parâmetros da mesma como pesos de filtros não paramétricos, de forma automática. A arquitetura da rede continha 189 entradas, 1 camada oculta contendo 5 neurônios e uma saída, conforme mostra a Figura 1.

Figura 1 – Arquitetura da rede neural



Fonte: Piva (2021).

Para a realização do treinamento do modelo, Piva (2021) utilizou o algoritmo resiliente *back-propagation with weight backtracking*, utilizando amostras das séries temporais referente ao período de 2012 a 2018 para treinar e estimar os parâmetros da rede neural. O autor avaliou a precisão através do Erro Percentual Médio Absoluto (EPMA), obtendo um desempenho de 1,929 de erro quadrático médio.

Na etapa de validação, Piva (2021) utilizou os dados do ano de 2019, no qual o modelo previu 3513 solicitações contra 3354 da série temporal real, resultando em um EPMA de 1,44%. O autor também observou que a despesa com educação infantil apresentou crescimento e, que houve decréscimo no orçamento destinado a melhorias e ampliação do atendimento da educação infantil.

O modelo de rede neural proposta por Piva (2021) produziu resultados próximos aos números reais, onde efeitos externos não podem ser previstos pelas séries temporais disponíveis atualmente, como a gestão da educação municipal sofre com influências externas. Tendo o trabalho abordado em apenas um município e a utilização de apenas uma arquitetura de rede neural, limitou a abrangência que o modelo poderia alcançar, mas isso não impede do modelo proposto ser aplicável em outros municípios. Piva (2021) sugere aplicar o modelo em outros municípios que tenham a mesma problemática e, que também sejam utilizados outros algoritmos e estratégias na projeção dos dados.

2.2 PREVISÃO DE MATRÍCULAS EM UMA IES UTILIZANDO SÉRIES TEMPORAIS E A SIMULAÇÃO DE MONTE CARLO

Pandolfi *et al.* (2014) disponibilizaram um modelo de previsão de matrículas para uma IES. Sendo que o foco era obter previsões de matrículas em uma IES usando séries temporais com simulação de diferentes cenários

Comentado [JM11]: Quais?

Comentado [JM12]: Trocar: "serão"

Por: "são"

Comentado [JM13]: Trocar: "proposta"

Por: "proposto"

Comentado [JM14]: Para suprir? Ou para prever?

Comentado [JM15]: Aqui não seria "no município de Camboriú"?

Comentado [JM16]: Excluir ","

Comentado [JM17]: Esse trecho não ficou muito claro, melhorar.

Comentado [JM18]: Esse trecho também não ficou claro, melhorar.

Comentado [JM19]: Excluir ","

Comentado [JM20]: Trocar: "para uma IES. Sendo que"

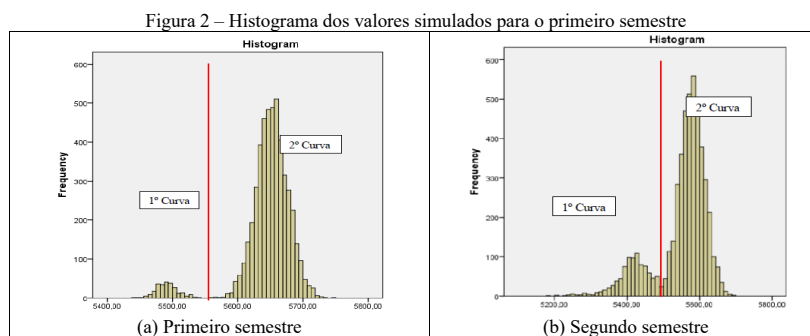
Por: "para uma IES, sendo que"

através do método Monte Carlo. Para isso, os autores utilizaram o histórico de matrícula dos últimos 14 anos, considerando também o crescimento do número de cursos e crise de mercado.

Segundo Pandolfi *et al.* (2014), coletaram os dados através de documentos internos da Instituição e dos enviados ao Ministério da Educação (MEC). A partir disso, os autores tentaram estabelecer o modelo para a projeções de matrículas. Pandolfi *et al.* (2014) optaram pelo modelo de decomposição multiplicativo para realizar a análise da série de tempo e a recomposição da tendência e sazonalidade. Para obter o último, os autores calcularam a média móvel dos últimos quatro períodos.

Pandolfi *et al.* (2014), aplicaram o teste de Kolmogorov-Smirnov, através do software Statistical Package for the Social Science (SPSS). Eles focaram em cursos da mesma área de conhecimento e com características semelhantes para determinar a média e o desvio padrão. Neste caso, utilizaram o Microsoft Excel para a geração de números aleatórios, formando 5000 valores para os cursos escolhidos. Por fim, Pandolfi *et al.* (2014) utilizaram o método de Monte Carlo para perfazer dois cenários para a Universidade, um cenário com e o outro cenário sem crise.

Pandolfi *et al.* (2014) simularam os cenários com e sem crise para obter o número de matrículas do próximo semestre. Neste processo, realizaram a soma da projeção de matrículas realizadas pelo modelo de séries com a soma do valor aleatório da matrícula do primeiro curso e valor aleatório de matrícula curso 2, subtraindo a parcela em função da crise e decréscimo de matrícula. Após estabelecer a previsão, os autores analisaram a distribuição resultante através de um histograma para ambos os semestres Figura 2, no qual no primeiro histograma (item a) têm-se o cenário com crise e no segundo (item b), o cenário sem crise.



Fonte: Pandolfi *et al.* (2014).

Pandolfi *et al.* (2014) observaram que no primeiro semestre a média de matrículas foi de 544,7 com desvio padrão de 42,46 matrículas. No segundo semestre, a média foi de 5550,37 matrículas com desvio padrão de 73,27. Os autores destacam que houve uma queda de matrículas no segundo semestre, ao qual respeitou a tendência histórica do modelo. Pandolfi *et al.* (2014) também observaram que os valores da média nominal e otimista estão próximos, justificando que a chance de crise econômica no primeiro semestre era baixa (5%). Já para o segundo semestre, apontou-se uma crise de 15%, mas que não houve uma diferença significativa nas médias nominal e otimista.

Pandolfi *et al.* (2014) concluem que a utilização de modelos de séries temporais alinhados com a simulação de Monte Carlo apresentou-se uma alternativa viável para obter previsões, mostrando-se adequada para diferentes cenários com função de acréscimos de novos cursos e decréscimo com uma possível crise. Os autores ressaltam que a utilização de Séries Temporais e a simulação de Monte Carlo para projeções de matrículas é totalmente aplicável. Segundo Pandolfi *et al.* (2014), os modelos desenvolvidos, facilitam a tomada de decisões dos gestores baseados em números, possibilitando a antecipação de algumas ações, como projeção de estrutura física, aumento de professores, receita e despesas. Por fim, os autores sugerem a realização de estudo do comportamento de matrículas por curso e sobre a perdas de alunos.

2.3 HIGHER EDUCATION INSTITUTION (HEI) ENROLLMENT FORECASTING USING DATA MINING TECHNIQUE

Cruz *et al.* (2020) tinham como objetivo prever o número de matrículas em um intervalo de 2019 a 2025 em uma instituição de ensino. Além disso, os autores tinham como intuito gerar informações para (i) fortalecer as políticas de admissão e retenção, (ii) tomar decisões estratégicas de gerenciamento de matrículas a longo prazo da universidade, (iii) desenvolver o plano anual de marketing e recrutamento, como também (iv) determinar os fatores internos e externos que afetam a queda e aumento da matrícula.

Comentado [JM21]: Trocar: "Segundo Pandolfi *et al.* (2014), coletaram"

Por: "Pandolfi *et al.* (2014) coletaram"

Comentado [JM22]: Excluir ", "

Comentado [JM23]: Excluir ", "

Comentado [JM24]: Excluir "cenário"

Comentado [JM25]: Esse trecho não ficou claro, melhorar.

Comentado [JM26]: Trocar: "Figura 2, no qual"

Por: "conforme pode ser observado na Figura 2", na qual

Comentado [JM27]: Trocar: "têm-se"

Por: "tem-se"

Comentado [JM28]: Esses valores estão certos?

Comentado [JM29]: Colocar em letras minúsculas.

Comentado [JM30]: Excluir ", "

Cruz *et al.* (2020) utilizaram os dados históricos do número total de matrículas de todos os cursos nos anos letivos de 2011-2019. Para isso, optaram pela aplicação do algoritmo AutoRegressive Integrated Moving Average (ARIMA) (p, d, q) para a previsão de matrículas. Além disso, os autores também empregaram a função de autocorrelação (ACF) e função parcial de autocorrelação (PACF), no qual a base da atribuição de p e q do modelo ARIMA é determinado por meio do gráfico de ACF e PACF. Cruz *et al.* (2020) justificam que o ARIMA é o modelo mais apropriado para a previsão com o menor valor de Critério de Informação Akaike (AIC).

Segundo Cruz *et al.* (2020), os dados obtidos foram imputados nas funções ACF e PACF visando descobrir os valores para o modelo ARIMA. A partir disso, os autores identificaram que o ACF e PACF estão decaindo, mas não de forma abrupta. A Tabela 1 apresenta os modelos ARIMA com menores valores AIC.

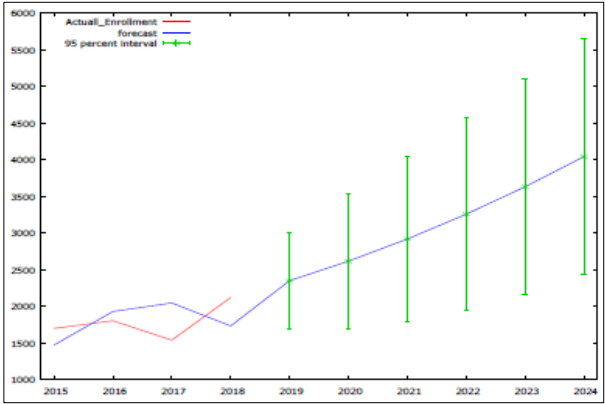
Tabela 1 – Modelos ARIMA com menor valor AIC

ARIMA Model	AIC
0,1,1	102.1462
** 0,2,1 **	94.68104

Fonte: Cruz *et al.* (2020).

A partir da Tabela 1, Cruz *et al.* (2020) observaram que o modelo ARIMA (0, 2, 1) é estatisticamente o modelo de previsão mais apropriado para estabelecer a taxa de matrícula. Neste cenário, segundo os autores, o modelo estabeleceu o menor valor AIC e a predição mais otimizada, com um intervalo de confiança de 95%. A Figura 3 apresenta a previsão das matrículas do campus CTU-Barili, usando o modelo ARIMA (0, 2, 1) com o seu intervalo de confiança.

Figura 3 – Previsão de número de matrículas para os anos seguintes



Fonte: Cruz *et al.* (2020).

Cruz *et al.* (2020) concluem que o modelo ARIMA (0, 2, 1) foi identificado como melhor modelo ARIMA (p, d, q) para previsão de matrículas na Universidade Tecnológica de Cebu. Os autores não chegam a apontar se o modelo pode ser aplicável para outras universidades, mas mostram que é possível aplicar o modelo ARIMA no contexto de prever matrículas em uma universidade para entender a tendência de matrículas.

Cruz *et al.* (2020) sugerem que em futuras pesquisas se considere na previsão, as contagens específicas de matrículas em instituições de **ensinos** para ter uma melhor compreensão da análise de tendência e extração de conhecimento. Além disso, também recomendam que sejam observadas técnicas e algoritmos de mineração de dados para uma maior extração de conhecimento.

3 PROPOSTA DO MÉTODO

Nesta seção **será** apresentada a justificativa para o desenvolvimento do trabalho, onde **é** apresentado os principais requisitos e metodologias de desenvolvimento que serão abordadas.

3.1 JUSTIFICATIVA

No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos. As linhas representam as características e as colunas os trabalhos.

Comentado [JM31]: A sigla apareceu antes e não foi colocada por extenso. Ajustar.

Comentado [JM32]: Trocar: "ensinos"

Por: "ensino"

Comentado [JM33]: Trocar: "será"

Por: "é"

Comentado [JM34]: Trocar: "é apresentado"

Por: "são apresentados"

Quadro 1 – Comparativo dos trabalhos correlatos

Trabalhos Correlatos Características	Piva (2021)	Pandolfi <i>et al.</i> (2014)	Cruz <i>et al.</i> (2020)
Objetivo	Criação de estratégias para suprir a demanda da educação infantil	Gestão da IES com base em um modelo de previsão	Estratégia de gerenciamento de matrículas a longo prazo da universidade
Principal algoritmo / técnica	Rede neural de múltiplas camadas	Modelos de série e simulação de Monte Carlo	ARIMA, ACF, PACF, AIC
Coleta de dados	Registros de fila de espera no sistema municipal	Documentos internos da Instituição e MEC	Dados históricos do número total de matrículas na IES

Fonte: elaborado pela autora.

A partir do Quadro 1 é possível observar que Pandolfi *et al.* (2014) e Cruz *et al.* (2020) elaboraram um modelo de previsão com o propósito de auxiliar na gestão e estratégias de gerenciamento de uma Instituição de Ensino Superior (IES). Por outro lado, Piva (2021) concentrou-se em atender à demanda da educação infantil em um determinado município.

Piva (2021) desenvolveu uma arquitetura de rede neural de múltiplas camadas visando prever a demanda na área da educação infantil. Já Pandolfi *et al.* (2014) utilizaram uma combinação de modelos de série com a simulação de Monte Carlo para a previsão de matrículas, concluindo que essa combinação foi adequada para realizar projeções em diferentes cenários. Em relação a Cruz *et al.* (2020), eles aplicaram o modelo ARIMA em conjunto com a técnica AIC para descobrir o melhor modelo para o ARIMA.

Pandolfi *et al.* (2014) e Cruz *et al.* (2020) fizeram uso dos dados históricos de matrículas de uma Instituição de Ensino Superior (IES), enquanto Piva (2021) baseou-se nos dados do registro de espera no sistema municipal.

Diante do contexto acima, este trabalho torna-se relevante pois disponibilizará uma compreensão mais aprofundada dos fatores que influenciam a demanda por instituições de ensino. Além disso, ele também possibilitará a identificação de desafios e oportunidades para melhorar a qualidade da educação, bem como o desenvolvimento de estratégias educacionais mais eficazes tanto para o governo quanto para as instituições privadas.

Espera-se que este estudo forneça estimativas confiáveis e precisas das matrículas futuras no ensino médio de Blumenau. Essas projeções podem ser valiosas para o planejamento educacional, ajudando os gestores escolares e os formuladores de políticas a tomarem decisões informadas sobre a alocação de recursos, a expansão de escolas existentes ou a construção de novas instituições de ensino. Além disso, espera-se que este estudo contribua para a compreensão do comportamento das matrículas no ensino médio em Blumenau ao longo do tempo e ofereça *insights* para pesquisas futuras relacionadas ao tema. É importante ressaltar que os resultados deste estudo estão sujeitos às limitações dos dados disponíveis e às suposições assumidas pelo modelo ARIMA.

3.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

O trabalho proposto deverá contemplar os seguintes Requisitos Funcionais (RF) e Requisitos Não Funcionais (RNF):

- permitir a entrada de dados históricos de matrículas do ensino médio de Blumenau (RF);
- utilizar técnicas de análise de dados para correlacionar, limpar e normalizar os dados (RF);
- permitir a parametrização do período de tempo (curto, médio e longo) (RF);
- realizar a análise dos dados utilizando o modelo ARIMA (RF);
- apresentar um intervalo de valores máximo e mínimo da previsão (RF);
- calcular as taxas de desempenho: média absoluta dos erros (MAE) e a raiz quadrada do erro quadrático médio (RMSE) (RF);
- utilizar as bibliotecas Scikit-Learn, pmdarima, statsmodels e a linguagem Python (RNF).

3.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- levantamento bibliográfico: pesquisar sobre sistema educacional, predição e séries temporais e trabalhos correlatos;
- entendimento dos dados: analisar as características e informações existentes na base de dados de matrículas do ensino médio da cidade de Blumenau;
- tratamento de dados: limpar a base de dados deixando apenas os registros consistentes (sem

Comentado [JM35]: Trocar:
"uma Instituição de Ensino Superior (IES)"

Por:
"IES"

Em toda essa seção

Comentado [JM36]: Trocar
"pois disponibilizará"

Por:
"pretende disponibilizar"

Comentado [JM37]: Aqui são instituições de ensino médio?

Comentado [JM38]: "Decisões informadas"... Que termo estranho. Acho que "decisões" é suficiente.

Comentado [JM39]: E contratação de professores? Lembra... Você colocou no objetivo principal.

Comentado [JM40]: Vai usar o modelo ARIMA? Não fica claro nos objetivos.

- informações faltantes), utilizando a linguagem Python e a biblioteca Pandas;
- definição das variáveis preditoras: definir, de forma *ad hoc*, quais variáveis possuem maior relevância e caracterizam a possibilidade de predição das matrículas ao longo do tempo;
 - pesquisa e escolha do algoritmo de predição: inicialmente será utilizado o modelo ARIMA. Porém, serão realizadas mais pesquisas para identificar outros modelos preditivos;
 - implementação: implementação do modelo levando em consideração os itens (b), (c), (d) e (e) utilizando a linguagem de programação Python a biblioteca Scikit-Learn, pmdarima, statsmodels;
 - análise do modelo: realizar testes a partir das bases de dados para verificar a eficiência e assertividade do modelo elaborado utilizando a validação cruzada, sensibilidade, erro médio absoluto (MAE) e o erro médio percentual absoluto (MAPE).

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 2 – Cronograma de atividades a serem realizadas

etapas / quinzenas	2023									
	jul.		ago.		set.		out.		nov.	
	1	2	1	2	1	2	1	2	1	2
levantamento bibliográfico										
entendimento dos dados										
tratamento de dados										
definição das variáveis preditoras										
pesquisa e escolha do algoritmo de predição										
implementação										
análise do modelo										

Fonte: elaborado pela autora.

4 REVISÃO BIBLIOGRÁFICA

Esta seção descreve brevemente sobre os assuntos que fundamentarão o estudo a ser realizado: Sistema educacional, predição e séries temporais.

Segundo Saviani (2010), sistemas educacionais são organizações complexas, instaladas em diferentes países, com o propósito de articular diversas atividades voltadas para a realização dos objetivos educativos das respectivas populações. Além disso, o sistema educacional é a unidade dos vários organismos e serviços educacionais intencionalmente reunidos de modo a formar um conjunto coerente que opera eficazmente no processo educativo da população a que se destina. Neste caso, de acordo com Saviani (2010), o Estado é o responsável da gestão e a regularização do sistema educativo, e deve garantir que o sistema inclua todas as pessoas.

Para Gooijer e Hyndman (2006), predição é o processo de estimar valores futuros de uma variável com base em dados históricos e outros fatores relevantes. Na análise de séries temporais, a predição é um dos principais objetivos, e envolve o uso de modelos estatísticos para gerar previsões precisas dos valores futuros de uma série temporal. De acordo com De Gooijer e Hyndman (2006), a predição em séries temporais é baseada na suposição de que padrões e comportamentos passados da série temporal continuarão no futuro, e que esses padrões podem ser identificados por meio de modelos estatísticos adequados.

De acordo com Hyndman e Athanasopoulos (2018), a análise de séries temporais é crucial para uma variedade de aplicações, desde finanças e economia até medicina e ciência dos materiais. Uma das principais razões para o uso de análise de séries temporais é que muitos sistemas dinâmicos são intrinsecamente dependentes do tempo, e o conhecimento de como as séries evoluem ao longo do tempo pode ajudar a tomar decisões informadas sobre esses sistemas. A análise de séries temporais envolve muitas técnicas, incluindo métodos de previsão baseados em modelos estatísticos, como o modelo ARIMA. Segundo os autores, ele é um modelo estatístico amplamente utilizado para previsão de séries temporais. Hyndman e Athanasopoulos (2018) também apontam que o ARIMA é um modelo que combina elementos de modelos de regressão (AR - Autoregressive) e modelos de média móvel (MA - Moving Average) com a capacidade de lidar com tendências e sazonalidade presentes em muitas séries temporais.

REFERÊNCIAS

- BODEWE, Robert-Jan. **Modelling and forecasting higher education Demand in the netherlands**. 2022. Teste (Mestrado em Gestão da Informação) - Curso de Pós-Graduação em Gestão da Informação, Universidade Tilburg, Holanda.
- CRONE, Sven F.; KOURENTZES, Nikolaos. Feature selection for time series prediction – a combined filter and wrapper approach for neural networks. *Neurocomputing*, [S.l.] v. 73, p. 1923-1936, Jun. 2010.
- CRUZ, Adeline P. Dela, *et al.* Higher Education Institution (HEI) Enrollment Forecasting Using Data Mining Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, [S.l.], v. 9, n. 2, Mar./Abr. 2020. Disponível em: <https://doi.org/10.30534/ijatcse/2020/179922020>. Acesso em: 24 mar. 2023.

Comentado [JM41]: Trocar: "Sistema"

Por: "sistema" com letra minúscula.

GARTNER. In: Gartner Glossary. [S.l.]: Gartner, 2023. Disponível em: <https://www.gartner.com/en/information-technology/glossary/predictive-modeling>. Acesso em: 23 abr. 2023.

GOOIJER, Jan G.; HYNDMAN, Rob J. 25 Years of Time Series Forecasting. **International Journal of Forecasting**, Jan. 2006. Disponível em: <https://www.robjhyndman.com/papers/ijf25.pdf>. Acesso em: 23 abr. 2023.

HYNDMAN, Rob J; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. 2. ed. [S. l.]: OTexts, 2018. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 23 abr. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. **InepData: Consulta Matrícula**. [S.l.], 2020. Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/inep-data/consulta-matricula>. Acesso em: 13 maio 2023.

_____. Censo da Educação Básica 2022: notas estatísticas. Brasília, DF: Inep, 2023. Disponível em: https://download.inep.gov.br/areas_de_atuacao/notas_estatisticas_censo_da_educacao_basica_2022.pdf. Acesso em: 13 maio 2023.

MACHADO, Eulália Nazaré C.; FALSARELLA, Ana Mari. Nova gestão pública, educação e gestão escolar. **Revista on line de Política e Gestão Educacional**, Araraquara, v. 24, n. 2, p. 372–89, Maio/Ago. 2020.

NEVES, Alan Vitor Coelho; RIGOTTI, José Irineu Rangel. Projeções de matrículas: análise e validação de métodos aplicados para o sistema educacional brasileiro. In: Anais do XXI Encontro Nacional de Estudos Populacionais, 2018, [S.l.]. **Anais...** [S.l.], 2018. Disponível em: <http://www.abep.org.br/xxiencontro/arquivos/R0203-1.pdf>. Acesso em: 13 maio 2023.

PANDOLFI, Cesar *et al.* Previsão de matrículas em uma IES utilizando séries temporais e a simulação de Monte Carlo. **Revista Global Manager** v. 14, n. 1, p. 130-146, 2014.

PIVA, Guilherme Offmaister. **Modelo de previsão de vagas de educação infantil de um município brasileiro**. 2021. Dissertação (Mestrado em Engenharia de Produção.) - Universidade de Caxias do Sul, [S. l.], 2021.

SAVIANI, Dermeval. **Gestrado UFMG: Sistemas Educacionais**. [S.l.], 2010. Disponível em: <https://gestrado.net.br/verbetes/sistemas-educacionais/>. Acesso em: 23 abr. 2023.

SOUZA, Ângelo Ricardo de. As condições de democratização da gestão da escola pública brasileira. **Ensaio: Avaliação e Políticas Públicas em Educação**, [S.l.], n. 103, p. 90-271, Jun. 2019. Disponível em: <https://www.scielo.br/j/ensaio/a/jFQH8xLn3TRvn964X7HCD6f/?lang=pt>. Acesso em: 23 abr. 2023.

FORMULÁRIO DE AVALIAÇÃO BCC – PROFESSOR AVALIADOR – PRÉ-PROJETO

Avaliador(a): **Joyce Martins**

Atenção: quando o avaliador marcar algum item como atende parcialmente ou não atende, deve obrigatoriamente indicar os motivos no texto, para que o aluno saiba o porquê da avaliação.

ASPECTOS AVALIADOS		Atende	atende parcialmente	não atende
ASPECTOS TÉCNICOS	1. INTRODUÇÃO O tema de pesquisa está devidamente contextualizado/delimitado?	x		
	O problema está claramente formulado? Deixar claro que é específico para o município de Blumenau (como destacado no título do trabalho).		x	
	2. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado? Deixar claro que é específico para o município de Blumenau (como destacado no título do trabalho). No objetivo principal, parece que também vai identificar a demanda por professores. É isso? Necessário rever/ajustar os objetivos específicos, requisitos, todo o texto.		x	
	Os objetivos específicos são coerentes com o objetivo principal? Quais modelos preditivos e quais técnicas de análise de séries temporais serão utilizadas/aplicadas? Quais são as variáveis relevantes para a predição da demanda por professores?		x	
	3. TRABALHOS CORRELATOS São apresentados trabalhos correlatos, bem como descritas as principais funcionalidades e os pontos fortes e fracos?	x		
	4. JUSTIFICATIVA Foi apresentado e discutido um quadro relacionando os trabalhos correlatos e suas principais funcionalidades com a proposta apresentada?	x		
	São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?		x	
	São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?		x	
	5. REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO Os requisitos funcionais e não funcionais foram claramente descritos? Verificar requisitos para identificar demanda por professores.		x	
	6. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?	x		
ASPECTOS METODOLÓGICOS	Os métodos, recursos e o cronograma estão devidamente apresentados e são compatíveis com a metodologia proposta? Incluir descrição para identificar demanda por professores.		x	
	7. REVISÃO BIBLIOGRÁFICA Os assuntos apresentados são suficientes e têm relação com o tema do TCC?	x		
	As referências contemplam adequadamente os assuntos abordados (são indicadas obras atualizadas e as mais importantes da área)?	x		
	8. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica? Poucos ajustes, apontados nos comentários.		x	
	A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)? Poucos ajustes, apontados nos comentários.		x	

Comentado [JM42]: Quais?

Comentado [JM43]: Quais?

CURSO DE CIÊNCIA DA COMPUTAÇÃO – TCC		
(X) PRÉ-PROJETO	() PROJETO	ANO/SEMESTRE: 2023/1

PREVISÃO DO NÚMERO DE MATRÍCULAS PARA O ENSINO MÉDIO DE BLUMENAU UTILIZANDO SÉRIES TEMPORAIS

Rossana Ariadna Schumann Dullius

Prof. Aurélio Faustino Hoppe – Orientador

5 INTRODUÇÃO

O contexto educacional do brasileiro ainda é desafiador, e a demanda por infraestrutura educacional é uma das principais questões a serem enfrentadas. Segundo Souza (2019), a precariedade da infraestrutura das escolas é um obstáculo para a qualidade da educação no país. A infraestrutura precária dificulta a inclusão de tecnologia e de atividades extracurriculares nas escolas, também a falta de investimento em reformas e manutenção de prédios escolares resulta em problemas estruturais, como salas superlotadas, falta de acessibilidade, pouca ventilação, dentre outros.

Segundo Machado e Falsarella (2020), a demanda escolar é uma preocupação constante de gestão pública no Brasil pois ela é responsável por planejar e gerenciar a oferta de vagas escolares, considerando as particularidades locais e as demandas dos estudantes. Ainda de acordo com os autores, neste caso, é importante que a gestão pública acompanhe a evolução da demanda escolar, para tanto é fundamental o investimento em políticas públicas que ampliem a oferta de vagas, como a construção de novas escolas e a ampliação das existentes, além das formas de incentivar a matrícula e permanência dos estudantes na escola.

No que se refere aos números totais de matrículas, segundo Instituto Nacional de Estudos e Pesquisas Educacionais (2022), o ensino básico como um todo, incluindo a educação de jovens e adultos, somou 35.760.081 matrículas em 2022, o que representa uma redução de 2,07% em relação aos dados preliminares de 2021. Esses dados são importantes não só para a gestão educacional, mas também para a distribuição de recursos do Fundo de Manutenção e Desenvolvimento da Educação Básica (Fundeb) e para a execução de programas na área da educação.

De acordo com Bodewe (2022), para realizar a projeção de matrículas com precisão, é necessário considerar diversas variáveis que influenciam o número de alunos matriculados, como número de nascimentos, migração de alunos entre escolas e evasão escolar. Além disso, é importante utilizar técnicas de análise de dados adequadas e contar com profissionais capacitados para realizar as análises. Dentre as técnicas de análise de séries temporais, destacam-se a análise de tendências e a análise de sazonalidade.

Neste contexto, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP) disponibiliza diversas ferramentas para análise de dados educacionais, incluindo o Consulta Matrícula, que apresenta o quantitativo de matrículas declaradas ao Censo Escolar anualmente e exibe um comparativo com o ano anterior (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS, 2020). Essa ferramenta pode ser útil para a realização de análises de séries temporais e para a projeção de matrículas no ensino médio. No entanto, é importante destacar que a predição de matrículas é uma atividade complexa que exige alto grau de precisão e confiabilidade. Por isso, é recomendável utilizar outras fontes de dados e técnicas de análise de dados para complementar a análise (NEVES; RIGOTTI, 2018).

De acordo com ~~Gartner~~ ~~-(GARTNER,~~ (2023), a predição é um processo analítico que utiliza modelos estatísticos e algoritmos de aprendizado de máquina para prever eventos futuros com base em dados históricos. Essa técnica é amplamente utilizada em diversos campos, incluindo negócios, finanças, medicina, engenharia, entre outros. Além disso, a predição também pode auxiliar no processo de projeções, permitindo que as organizações prevejam demanda futura e tomem decisões estratégicas.

A partir deste contexto, surge a seguinte pergunta de pesquisa: a utilização de modelos preditivos em conjunto com séries temporais pode auxiliar nas projeções da demanda municipal de escolas e professores do ensino médio?

5.1 OBJETIVOS

O objetivo deste trabalho é analisar como a utilização de modelos preditivos em conjunto com séries temporais pode contribuir para projeções precisas da demanda municipal de escolas e professores no ensino médio.

Os objetivos específicos são:

- h) coletar e analisar os dados históricos de matrículas no ensino médio em Blumenau;
- i) identificar as variáveis relevantes para a predição da demanda de matrículas, como o número de alunos ingressantes no ensino fundamental e o índice de natalidade;

- j) aplicar técnicas de análise de séries temporais para estimar o número de matrículas no ensino médio considerando cenários de curto, médio e longo prazo;
- k) avaliar a precisão e confiabilidade do modelo desenvolvido.

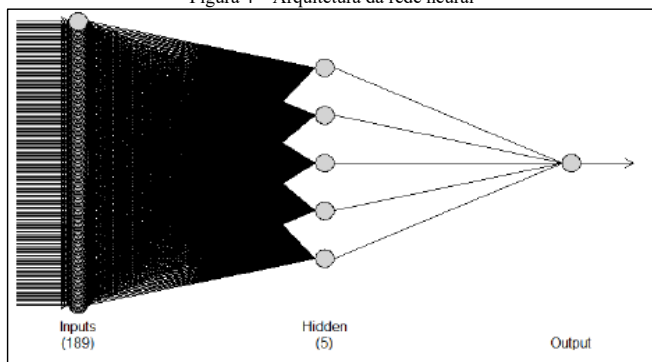
6 TRABALHOS CORRELATOS

Nesta seção serão apresentados três trabalhos que possuem semelhanças com o trabalho proposto. Na subseção 2.1 encontra-se o trabalho de Piva (2021), ao qual utiliza rede neural a fim de criar um modelo de previsão para suprir a demanda escolar. Na subseção 2.2 descreve-se a aplicação do modelo de Monte Carlo para previsão de matrículas para gerar estratégias a uma Instituição de Ensino Superior (IES) (PANDOLFI *et al.*, 2014). Por fim, na subseção 2.3 detalha-se o trabalho de Cruz *et al.* (2020) que apresentam o uso do algoritmo ARIMA para a realização de previsão de matrículas em uma IES.

6.1 MODELO DE PREVISÃO DE VAGAS DE EDUCAÇÃO INFANTIL DE UM MUNICÍPIO BRASILEIRO

Piva (2021) desenvolveu um modelo utilizando a arquitetura de rede neural multicamadas visando prever a demanda da educação infantil em um município brasileiro de Camboriú. Para isso, o autor coletou informações dos registros para a fila de espera do sistema municipal. Posteriormente, formulou a série temporal com os inscritos da fila de espera. Para implementação, Piva (2021) utilizou a biblioteca *nnfor* da linguagem R. Os parâmetros utilizados foram definidos utilizando o método proposto por Crone e Kourentzes (2010), onde estima-se os parâmetros da mesma como pesos de filtros não paramétricos, de forma automática. A arquitetura da rede continha 189 entradas, 1 camada oculta contendo 5 neurônios e uma saída, conforme mostra a Figura 1.

Figura 4 – Arquitetura da rede neural



Fonte: Piva (2021).

Para a realização do treinamento do modelo, Piva (2021) utilizou o algoritmo resiliente *back-propagation with weight backtrain*, utilizando amostras das séries temporais referente ao período de 2012 a 2018 para treinar e estimar os parâmetros da rede neural. O autor avaliou a precisão através do Erro Percentual Médio Absoluto (EPMA), obtendo um desempenho de 1,929 de erro quadrático médio.

Na etapa de validação, Piva (2021) utilizou os dados do ano de 2019, no qual o modelo previu 3513 solicitações contra 3354 da série temporal real, resultando em um EPMA de 1,44%. O autor também observou que a despesa com educação infantil apresentou crescimento e, que houve decréscimo no orçamento destinado a melhorias e ampliação do atendimento da educação infantil.

O modelo de rede neural proposta por Piva (2021) produziu resultados próximos aos números reais, onde efeitos externos não podem ser previstos pelas séries temporais disponíveis atualmente, como a gestão da educação municipal sofre com influências externas. Tendo o trabalho abordado em apenas um município e a utilizam de apenas uma arquitetura de rede neural, limitou a abrangência que o modelo poderia alcançar, mas isso não impede do modelo proposto ser aplicável em outros municípios. Piva (2021) sugere aplicar o modelo em outros municípios que tenham a mesma problemática e, que também sejam utilizados outros algoritmos e estratégias na projeção dos dados.

6.2 PREVISÃO DE MATRÍCULAS EM UMA IES UTILIZANDO SÉRIES TEMPORAIS E A SIMULAÇÃO DE MONTE CARLO

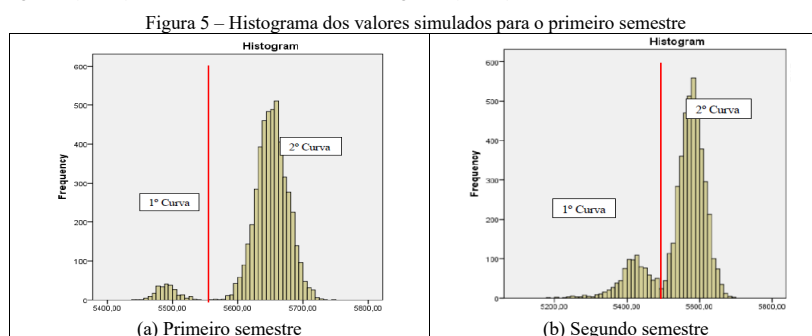
Pandolfi *et al.* (2014) disponibilizaram um modelo de previsão de matrículas para uma IES. Sendo que o foco era obter previsões de matrículas em uma IES usando séries temporais com simulação de diferentes cenários

através do método Monte Carlo. Para isso, os autores utilizaram o histórico de matrícula dos últimos 14 anos, considerando também o crescimento do número de cursos e crise de mercado.

Segundo Pandolfi *et al.* (2014), coletaram os dados através de documentos internos da Instituição e dos enviados ao Ministério da Educação (MEC). A partir disso, os autores tentaram estabelecer o modelo para a projeções de matrículas. Pandolfi *et al.* (2014) optaram pelo modelo de decomposição multiplicativo para realizar a análise da série de tempo e a recomposição da tendência e sazonalidade. Para obter o último, os autores calcularam a média móvel dos últimos quatro períodos.

Pandolfi *et al.* (2014), aplicaram o teste de Kolmogorov-Smirnov, através do software Statistical Package for the Social Science (SPSS). Eles focaram em cursos da mesma área de conhecimento e com características semelhantes para determinar a média e o desvio padrão. Neste caso, utilizaram o Microsoft Excel para a geração de números aleatórios, formando 5000 valores para os cursos escolhidos. Por fim, Pandolfi *et al.* (2014) utilizaram o método de Monte Carlo para perfazer dois cenários para a Universidade, um cenário com e o outro cenário sem crise.

Pandolfi *et al.* (2014) simularam os cenários com e sem crise para obter o número de matrículas do próximo semestre. Neste processo, realizaram a soma da projeção de matrículas realizadas pelo modelo de séries com a soma do valor aleatório da matrícula do primeiro curso e valor aleatório de matrícula curso 2, subtraindo a parcela em função da crise e decréscimo de matrícula. Após estabelecer a previsão, os autores analisaram a distribuição resultante através de um histograma para ambos os semestres (Figura 2), no qual no primeiro histograma (item a) têm-se o cenário com crise e no segundo (item b), o cenário sem crise.



Fonte: Pandolfi *et al.* (2014).

Pandolfi *et al.* (2014) observaram que no primeiro semestre a média de matrículas foi de 544,7 com desvio padrão de 42,46 matrículas. No segundo semestre, a média foi de 5550,37 matrículas com desvio padrão de 73,27. Os autores destacam que houve uma queda de matrículas no segundo semestre, ao qual respeitou a tendência histórica do modelo. Pandolfi *et al.* (2014) também observaram que os valores da média nominal e otimista estão próximos, justificando que a chance de crise econômica no primeiro semestre era baixa (5%). Já para o segundo semestre, apontou-se uma crise de 15%, mas que não houve uma diferença significativa nas médias nominal e otimista.

Pandolfi *et al.* (2014) concluem que a utilização de modelos de séries temporais alinhados com a simulação de Monte Carlo apresentou-se uma alternativa viável para obter previsões, mostrando-se adequada para diferentes cenários com função de acréscimos de novos cursos e decréscimo com uma possível crise. Os autores ressaltam que a utilização de Séries Temporais e a simulação de Monte Carlo para projeções de matrículas é totalmente aplicável. Segundo Pandolfi *et al.* (2014), os modelos desenvolvidos, facilitam a tomada de decisões dos gestores baseados em números, possibilitando a antecipação de algumas ações, como projeção de estrutura física, aumento de professores, receita e despesas. Por fim, os autores sugerem a realização de estudo do comportamento de matrículas por curso e sobre a perdas de alunos.

6.3 HIGHER EDUCATION INSTITUTION (HEI) ENROLLMENT FORECASTING USING DATA MINING TECHNIQUE

Cruz *et al.* (2020) tinham como objetivo prever o número de matrículas em um intervalo de 2019 a 2025 em uma instituição de ensino. Além disso, os autores tinham como intuito gerar informações para (i) fortalecer as políticas de admissão e retenção, (ii) tomar decisões estratégicas de gerenciamento de matrículas a longo prazo da universidade, (iii) desenvolver o plano anual de marketing e recrutamento, como também (iv) determinar os fatores internos e externos que afetam a queda e aumento da matrícula.

Cruz *et al.* (2020) utilizaram os dados históricos do número total de matrículas de todos os cursos nos anos letivos de 2011-2019. Para isso, optaram pela aplicação do algoritmo AutoRegressive Integrated Moving Average (ARIMA) (p, d, q) para a previsão de matrículas. Além disso, os autores também empregaram a função de autocorrelação (ACF) e função parcial de autocorrelação (PACF), no qual a base da atribuição de p e q do modelo ARIMA é determinado por meio do gráfico de ACF e PACF. Cruz *et al.* (2020) justificam que o ARIMA é o modelo mais apropriado para a previsão com o menor valor de Critério de Informação Akaike (AIC).

Segundo Cruz *et al.* (2020), os dados obtidos foram imputados nas funções ACF e PACF visando descobrir os valores para o modelo ARIMA. A partir disso, os autores identificaram que o ACF e PACF estão decaindo, mas não de forma abrupta. A Tabela 1 apresenta os modelos ARIMA com menores valores AIC.

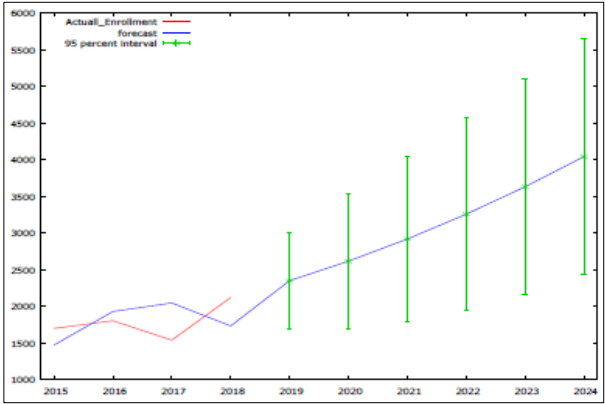
Tabela 2 – Modelos ARIMA com menor valor AIC

ARIMA Model	AIC
0,1,1	102.1462
** 0,2,1 **	94.68104

Fonte: Cruz *et al.* (2020).

A partir da Tabela 1, Cruz *et al.* (2020) observaram que o modelo ARIMA (0, 2, 1) é estatisticamente o modelo de previsão mais apropriado para estabelecer a taxa de matrícula. Neste cenário, segundo os autores, o modelo estabeleceu o menor valor AIC e a predição mais otimizada, com um intervalo de confiança de 95%. A Figura 3 apresenta a previsão das matrículas do campus CTU-Barili, usando o modelo ARIMA (0, 2, 1) com o seu intervalo de confiança.

Figura 6 – Previsão de número de matrículas para os anos seguintes



Fonte: Cruz *et al.* (2020).

Cruz *et al.* (2020) concluem que o modelo ARIMA (0, 2, 1) foi identificado como melhor modelo ARIMA (p, d, q) para previsão de matrículas na Universidade Tecnológica de Cebu. Os autores não chegam a apontar se o modelo pode ser aplicável para outras universidades, mas mostram que é possível aplicar o modelo ARIMA no contexto de prever matrículas em uma universidade para entender a tendência de matrículas.

Cruz *et al.* (2020) sugerem que em futuras pesquisas se considere na previsão, as contagens específicas de matrículas em instituições de ensino para ter uma melhor compreensão da análise de tendência e extração de conhecimento. Além disso, também recomendam que sejam observadas técnicas e algoritmos de mineração de dados para uma maior extração de conhecimento.

7 PROPOSTA DO MÉTODO

Nesta seção será apresentada a justificativa para o desenvolvimento do trabalho, onde é apresentado os principais requisitos e metodologias de desenvolvimento que serão abordadas.

7.1 JUSTIFICATIVA

No Quadro 1 é apresentado um comparativo entre os trabalhos correlatos. As linhas representam as características e as colunas os trabalhos.

Quadro 3 – Comparativo dos trabalhos correlatos

Trabalhos Correlatos Características	Piva (2021)	Pandolfi <i>et al.</i> (2014)	Cruz <i>et al.</i> (2020)
Objetivo	Criação de estratégias para suprir a demanda da educação infantil	Gestão da IES com base em um modelo de previsão	Estratégia de gerenciamento de matrículas a longo prazo da universidade
Principal algoritmo / técnica	Rede neural de múltiplas camadas	Modelos de série e simulação de Monte Carlo	ARIMA, ACF, PACF, AIC
Coleta de dados	Registros de fila de espera no sistema municipal	Documentos internos da Instituição e MEC	Dados históricos do número total de matrículas na IES

Fonte: elaborado pela autora.

A partir do Quadro 1 é possível observar que Pandolfi *et al.* (2014) e Cruz *et al.* (2020) elaboraram um modelo de previsão com o propósito de auxiliar na gestão e estratégias de gerenciamento de uma Instituição de Ensino Superior (IES). Por outro lado, Piva (2021) concentrou-se em atender à demanda da educação infantil em um determinado município.

Piva (2021) desenvolveu uma arquitetura de rede neural de múltiplas camadas visando prever a demanda na área da educação infantil. Já Pandolfi *et al.* (2014) utilizaram uma combinação de modelos de série com a simulação de Monte Carlo para a previsão de matrículas, concluindo que essa combinação foi adequada para realizar projeções em diferentes cenários. Em relação a Cruz *et al.* (2020), eles aplicaram o modelo ARIMA em conjunto com a técnica AIC para descobrir o melhor modelo para o ARIMA.

Pandolfi *et al.* (2014) e Cruz *et al.* (2020) fizeram uso dos dados históricos de matrículas de uma Instituição de Ensino Superior (IES), enquanto Piva (2021) baseou-se nos dados do registro de espera no sistema municipal.

Diante do contexto acima, este trabalho torna-se relevante pois disponibilizará uma compreensão mais aprofundada dos fatores que influenciam a demanda por instituições de ensino. Além disso, ele também possibilitará a identificação de desafios e oportunidades para melhorar a qualidade da educação, bem como o desenvolvimento de estratégias educacionais mais eficazes tanto para o governo quanto para as instituições privadas.

Espera-se que este estudo forneça estimativas confiáveis e precisas das matrículas futuras no ensino médio de Blumenau. Essas projeções podem ser valiosas para o planejamento educacional, ajudando os gestores escolares e os formuladores de políticas a tomarem decisões informadas sobre a alocação de recursos, a expansão de escolas existentes ou a construção de novas instituições de ensino. Além disso, espera-se que este estudo contribua para a compreensão do comportamento das matrículas no ensino médio em Blumenau ao longo do tempo e ofereça *insights* para pesquisas futuras relacionadas ao tema. É importante ressaltar que os resultados deste estudo estão sujeitos às limitações dos dados disponíveis e às suposições assumidas pelo modelo ARIMA.

7.2 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

O trabalho proposto deverá contemplar os seguintes Requisitos Funcionais (RF) e Requisitos Não Funcionais (RNF):

- l) permitir a entrada de dados históricos de matrículas do ensino médio de Blumenau (RF);
- m) utilizar técnicas de análise de dados para correlacionar, limpar e normalizar os dados (RF);
- n) permitir a parametrização do período de tempo (curto, médio e longo) (RF);
- o) realizar a análise dos dados utilizando o modelo ARIMA (RF);
- p) apresentar um intervalo de valores máximo e mínimo da previsão (RF);
- q) calcular as taxas de desempenho: média absoluta dos erros (MAE) e a raiz quadrada do erro quadrático médio (RMSE) (RF);
- r) utilizar as bibliotecas Scikit-Learn, pmdarima, statsmodels e a linguagem Python (RNF).

7.3 METODOLOGIA

O trabalho será desenvolvido observando as seguintes etapas:

- s) levantamento bibliográfico: pesquisar sobre sistema educacional, predição e séries temporais e trabalhos correlatos;
- t) entendimento dos dados: analisar as características e informações existentes na base de dados de matrículas do ensino médio da cidade de Blumenau;
- u) tratamento de dados: limpar a base de dados deixando apenas os registros consistentes (sem

- informações faltantes), utilizando a linguagem Python e a biblioteca Pandas;
- v) definição das variáveis preditoras: definir, de forma *ad hoc*, quais variáveis possuem maior relevância e caracterizam a possibilidade de predição das matrículas ao longo do tempo;
 - w) pesquisa e escolha do algoritmo de predição: inicialmente será utilizado o modelo ARIMA. Porém, serão realizadas mais pesquisas para identificar outros modelos preditivos;
 - x) implementação: implementação do modelo levando em consideração os itens (b), (c), (d) e (e) utilizando a linguagem de programação Python a biblioteca Scikit-Learn, pmdarima, statsmodels;
 - y) análise do modelo: realizar testes a partir das bases de dados para verificar a eficiência e assertividade do modelo elaborado utilizando a validação cruzada, sensibilidade, erro médio absoluto (MAE) e o erro médio percentual absoluto (MAPE).

As etapas serão realizadas nos períodos relacionados no Quadro 2.

Quadro 4 – Cronograma de atividades a serem realizadas

etapas / quinzenas	2023									
	jul.		ago.		set.		out.		nov.	
	1	2	1	2	1	2	1	2	1	2
levantamento bibliográfico										
entendimento dos dados										
tratamento de dados										
definição das variáveis preditoras										
pesquisa e escolha do algoritmo de predição										
implementação										
análise do modelo										

Fonte: elaborado pela autora.

8 REVISÃO BIBLIOGRÁFICA

Esta seção descreve brevemente sobre os assuntos que fundamentarão o estudo a ser realizado: Sistema educacional, predição e séries temporais.

Segundo Saviani (2010), sistemas educacionais são organizações complexas, instaladas em diferentes países, com o propósito de articular diversas atividades voltadas para a realização dos objetivos educativos das respectivas populações. Além disso, o sistema educacional é a unidade dos vários organismos e serviços educacionais intencionalmente reunidos de modo a formar um conjunto coerente que opera eficazmente no processo educativo da população a que se destina. Neste caso, de acordo com Saviani (2010), o Estado é o responsável da gestão e a regularização do sistema educativo, e deve garantir que o sistema inclua todas as pessoas.

Para Gooijer e Hyndman (2006), predição é o processo de estimar valores futuros de uma variável com base em dados históricos e outros fatores relevantes. Na análise de séries temporais, a predição é um dos principais objetivos, e envolve o uso de modelos estatísticos para gerar previsões precisas dos valores futuros de uma série temporal. De acordo com De-Gooijer e Hyndman (2006), a predição em séries temporais é baseada na suposição de que padrões e comportamentos passados da série temporal continuarão no futuro, e que esses padrões podem ser identificados por meio de modelos estatísticos adequados.

De acordo com Hyndman e Athanasopoulos (2018), a análise de séries temporais é crucial para uma variedade de aplicações, desde finanças e economia até medicina e ciência dos materiais. Uma das principais razões para o uso de análise de séries temporais é que muitos sistemas dinâmicos são intrinsecamente dependentes do tempo, e o conhecimento de como as séries evoluem ao longo do tempo pode ajudar a tomar decisões informadas sobre esses sistemas. A análise de séries temporais envolve muitas técnicas, incluindo métodos de previsão baseados em modelos estatísticos, como o modelo ARIMA. Segundo os autores, ele é um modelo estatístico amplamente utilizado para previsão de séries temporais. Hyndman e Athanasopoulos (2018) também apontam que o ARIMA é um modelo que combina elementos de modelos de regressão (AR - Autoregressive) e modelos de média móvel (MA - Moving Average) com a capacidade de lidar com tendências e sazonalidade presentes em muitas séries temporais.

REFERÊNCIAS

- BODEWE, Robert-Jan. **Modelling and forecasting higher education Demand in the netherlands**. 2022. Teste (Mestrado em Gestão da Informação) - Curso de Pós-Graduação em Gestão da Informação, Universidade Tilburg, Holanda.
- CRONE, Sven F.; KOURENTZES, Nikolaos. Feature selection for time series prediction – a combined filter and wrapper approach for neural networks. *Neurocomputing*, [S.l.] v. 73, p. 1923-1936, Jun. 2010.
- CRUZ, Adeline P. Dela, *et al.* Higher Education Institution (HEI) Enrollment Forecasting Using Data Mining Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, [S.l.], v. 9, n. 2, Mar./Abr. 2020. Disponível em: <https://doi.org/10.30534/ijatcse/2020/179922020>. Acesso em: 24 mar. 2023.

GARTNER. In: Gartner Glossary. [S.l.]: Gartner, 2023. Disponível em: <https://www.gartner.com/en/information-technology/glossary/predictive-modeling>. Acesso em: 23 abr. 2023.

GOOIJER, Jan G.; HYNDMAN, Rob J. 25 Years of Time Series Forecasting. **International Journal of Forecasting**, Jan. 2006. Disponível em: <https://www.robjhyndman.com/papers/ijf25.pdf>. Acesso em: 23 abr. 2023.

HYNDMAN, Rob J; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. 2. ed. [S. l.]: OTexts, 2018. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 23 abr. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. **InepData: Consulta Matrícula**. [S.l.], 2020. Disponível em: <https://www.gov.br/inep/pt-br/acao-a-informacao/dados-abertos/inep-data/consulta-matricula>. Acesso em: 13 maio 2023.

_____. Censo da Educação Básica 2022: notas estatísticas. Brasília, DF: Inep, 2023. Disponível em: https://download.inep.gov.br/areas_de_atuacao/notas_estatisticas_censo_da_educacao_basica_2022.pdf. Acesso em: 13 maio 2023.

MACHADO, Eulália Nazaré C.; FALSARELLA, Ana Mari. Nova gestão pública, educação e gestão escolar. **Revista on line de Política e Gestão Educacional**, Araraquara, v. 24, n. 2, p. 372–89, Maio/Ago. 2020.

NEVES, Alan Vitor Coelho; RIGOTTI, José Irineu Rangel. Projeções de matrículas: análise e validação de métodos aplicados para o sistema educacional brasileiro. In: Anais do XXI Encontro Nacional de Estudos Populacionais, 2018, [S.l.]. **Anais...** [S.l.], 2018. Disponível em: <http://www.abep.org.br/xxiencontro/arquivos/R0203-1.pdf>. Acesso em: 13 maio 2023.

PANDOLFI, Cesar *et al.* Previsão de matrículas em uma IES utilizando séries temporais e a simulação de Monte Carlo. **Revista Global Manager** v. 14, n. 1, p. 130-146, 2014.

PIVA, Guilherme Offmaister. **Modelo de previsão de vagas de educação infantil de um município brasileiro**. 2021. Dissertação (Mestrado em Engenharia de Produção.) - Universidade de Caxias do Sul, [S. l.], 2021.

SAVIANI, Dermeval. **Gestrado UFMG: Sistemas Educacionais**. [S.l.], 2010. Disponível em: <https://gestrado.net.br/verbetes/sistemas-educacionais/>. Acesso em: 23 abr. 2023.

SOUZA, Ângelo Ricardo de. As condições de democratização da gestão da escola pública brasileira. **Ensaio: Avaliação e Políticas Públicas em Educação**, [S.l.], n. 103, p. 90-271, Jun. 2019. Disponível em: <https://www.scielo.br/j/ensaio/a/jFQH8xLn3TRvn964X7HCD6f/?lang=pt>. Acesso em: 23 abr. 2023.

FORMULÁRIO DE AVALIAÇÃO BCC – PROFESSOR TCC I – PRÉ-PROJETO

Avaliador(a): Dalton Solano dos Reis

ASPECTOS AVALIADOS		atende	atende parcialmente	não atende
ASPECTOS TÉCNICOS	9. INTRODUÇÃO O tema de pesquisa está devidamente contextualizado/delimitado?	X		
	O problema está claramente formulado?	X		
	10. OBJETIVOS O objetivo principal está claramente definido e é passível de ser alcançado?	X		
	Os objetivos específicos são coerentes com o objetivo principal?	X		
	11. JUSTIFICATIVA São apresentados argumentos científicos, técnicos ou metodológicos que justificam a proposta?	X		
	São apresentadas as contribuições teóricas, práticas ou sociais que justificam a proposta?	X		
	12. METODOLOGIA Foram relacionadas todas as etapas necessárias para o desenvolvimento do TCC?	X		
	Os métodos, recursos e o cronograma estão devidamente apresentados?	X		
ASPECTOS METODOLÓGICOS	13. REVISÃO BIBLIOGRÁFICA Os assuntos apresentados são suficientes e têm relação com o tema do TCC?	X		
	14. LINGUAGEM USADA (redação) O texto completo é coerente e redigido corretamente em língua portuguesa, usando linguagem formal/científica?	X		
	A exposição do assunto é ordenada (as ideias estão bem encadeadas e a linguagem utilizada é clara)?	X		
	15. ORGANIZAÇÃO E APRESENTAÇÃO GRÁFICA DO TEXTO A organização e apresentação dos capítulos, seções, subseções e parágrafos estão de acordo com o modelo estabelecido?	X		
	16. ILUSTRAÇÕES (figuras, quadros, tabelas) As ilustrações são legíveis e obedecem às normas da ABNT?	X		
	17. REFERÊNCIAS E CITAÇÕES As referências obedecem às normas da ABNT?	X		
	As citações obedecem às normas da ABNT?	X		
	Todos os documentos citados foram referenciados e vice-versa, isto é, as citações e referências são consistentes?	X		

8.1.1



UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
DISCIPLINA: TRABALHO DE CONCLUSÃO DE CURSO I
CURSO: CIÊNCIA DA COMPUTAÇÃO - BCC

ATA DA DEFESA: BANCA DO PRÉ-PROJETO

Venho, por meio deste, manifestar minha avaliação sobre a **apresentação** do Pré-Projeto de TCC realizado pelo(a) acadêmico(a), **_ Rossana Ariadna Schumann Dullius _** no **PRIMEIRO SEMESTRE DE 2023**, com o título **_ PREVISÃO DO NÚMERO DE MATRÍCULAS PARA O ENSINO MÉDIO DE BLUMENAU UTILIZANDO SÉRIES TEMPORAIS _**, sob orientação do prof(a). **_ Aurélio Faustino Hoppe _**.

A referida apresentação obteve a seguinte nota:

Componente da Banca	Nota (de 0 a 10)
Professor(a) Avaliador(a): _ Joyce Martins _	9,0

ATENÇÃO. A nota acima se refere somente a apresentação do pré-projeto e vai ser repassada para o aluno (orientando). Favor preencher os campos acima e enviar por e-mail ao professor de TCC1 (dalton@furb.br). Não passar o arquivo com as anotações da revisão já enviado ao professor de TCC1 para o orientando e nem para o professor orientador. Após o professor de TCC1 receber esta ata preenchida, o professor de TCC1 vai disponibilizar para o orientando/orientador os arquivos com as revisões. Caso julgue necessário fazer mais alguma consideração relacionada ao pré-projeto ou a defesa, favor usar o espaço abaixo.

Observações da apresentação: Para TCC 2, utilizar melhor o tempo para a apresentação.



ATA DA DEFESA: BANCA DO PRÉ-PROJETO

Venho, por meio deste, manifestar minha avaliação sobre a **apresentação** do Pré-Projeto de TCC realizado pelo(a) acadêmico(a), **_ Rossana Ariadna Schumann Dullius _** no **PRIMEIRO SEMESTRE DE 2023**, com o título **_ PREVISÃO DO NÚMERO DE MATRÍCULAS PARA O ENSINO MÉDIO DE BLUMENAU UTILIZANDO SÉRIES TEMPORAIS _**.

A referida apresentação obteve a seguinte nota:

Componente da Banca	Nota (de 0 a 10)
Professor(a) Avaliador(a): _ Aurélio Faustino Hoppe _	8,5

A apresentação aconteceu em 06 / 06 / 2023 na sala de reunião virtual do MS-Teams, tendo início às 18 : 03 hs e foi encerrada às 18 : 30 hs.

ATENÇÃO. A nota acima se refere somente a apresentação do pré-projeto e vai ser repassada para o aluno (orientando). Favor preencher os campos acima e enviar por e-mail ao professor de TCC1 (dalton@furb.br). Não passar o arquivo com as anotações da revisão já enviado ao professor de TCC1 para o orientando e nem para o professor orientador. Após o professor de TCC1 receber esta ata preenchida, o professor de TCC1 vai disponibilizar para o orientando/orientador os arquivos com as revisões. Caso julgue necessário fazer mais alguma consideração relacionada ao pré-projeto ou a defesa, favor usar o espaço abaixo.

Observações da apresentação: