Uma Análise de Modelos de Séries Temporais para Predição de Evasão Discente: Estudo de Caso do IFCE

Fernando Wagner Brito Hortêncio Filho [1]
Tiago Silva Vinuto [2]
Bruno de Carvalho Leal [3]

RESUMO

Ao longo da última década os institutos federais passaram por uma fase de forte expansão e interiorização, aumentando o acesso à educação técnica e profissionalizante. Em paralelo ao aumento do acesso a educação técnica e profissional o problema da evasão discente ganhou destaque, chamando a atenção dos gestores das instituições educacionais. Várias ações vem sendo elaboradas visando a diminuição dos índice de alunos evadidos, fortalecendo políticas de combate a evasão. Neste contexto, técnicas estatísticas vem sendo utilizadas no sentido de auxiliar o entendimento e compreensão dos dados relacionados a evasão, provendo suporte para uma melhor elaboração de planos e tomadas de decisão. O presente trabalho tem como objetivo analisar a viabilidade e o potencial de técnicas de construção de modelos de séries temporais para predição das taxas de alunos candidatos a evasão. Utilizou-se dados de nove campi de primeira geração do Instituto Federal do Ceará, compreendido entre os períodos de 2009 à 2017 para construção dos modelos (treinamento), tendo como alvo a predição de possíveis alunos evadidos que ingressaram no ano de 2018 (teste). Os resultados foram considerados promissores, sinalizando grande potencial em relação às técnicas utilizadas.

Palavras-chave: Evasão discente, Predição, Séries temporais, Instituto Federal.

INTRODUÇÃO

Nos últimos anos, ferramentas computacionais e métodos estatísticos vêm sendo utilizados com a finalidade de elucidar questões inerentes às várias esferas da sociedade. Estes mecanismos são aplicados como meio de resolver tarefas, otimizar processos e, por conseguinte, agregar valor frente a outras áreas de estudo. Na saúde, a utilização de tecnologias para provimento da telemedicina e a construção de prontuários eletrônicos vêm ganhando bastante atenção de profissionais e pesquisadores. Na economia e na administração, métodos estatísticos aliados às ferramentas computacionais vêm se destacando na construção e implantação de sistemas de apoio à tomada de decisão, bem

[1] Docente de Informática do Instituto Federal de Brasília - IFB, fernand.filho@ifb.edu.br;
[2] Doutorando em Computação da Universidade Federal do Ceará - UFC, tiagosv@lia.ufc.br;
[3] Docente de Informática do Instituto Federal do Piauí - IFPI, brunoleal@ifpi.edu.br;

como modelos econométricos para determinação e previsão de receitas e despesas em uma determinada instituição.

Na esfera educacional, a evolução da informática tornou possível a criação de soluções visando facilitar o ensino de componentes curriculares e viabilizar a educação a distância (EaD), como o desenvolvimento dos AVAs (Ambientes Virtuais de Aprendizagem) e os objetos de aprendizagem, que visam facilitar o processo de ensino e aprendizagem entre docente e aluno. No entanto, os benefícios nesta área não se limitam apenas às questões relacionadas ao processo de ensino e aprendizagem. Diversas ferramentas e algoritmos voltados para uso em áreas como acessibilidade e gestão educacional vêm surgindo.

Atualmente, um dos maiores problemas educacionais enfrentados pelas instituições de ensino é a evasão discente, que é caracterizada pelo desligamento do aluno antes da finalização de todas as etapas necessárias para a conclusão do curso. Via de regra, a evasão é um problema de difícil tratamento podendo ser motivada por causas variadas, fazendo-se necessária o entendimento deste fenômeno pela instituição de ensino. O uso de técnicas computacionais podem auxiliar a instituição a lidar melhor com o fenômeno da evasão, provendo suporte a medidas proativas, e não somente reativas de combate.

O presente trabalho tem como objetivo elaborar uma análise de modelos de séries temporais, utilizando como estudo de caso os números de evasão dos campi de primeira geração do IFCE. Além de uma prévia análise exploratória, são utilizadas técnicas para a construção de modelos temporais e discutido o potencial das mesmas na predição das taxas periódicas de evasão, de modo a auxiliar no diagnóstico e combate da evasão discente.

REFERENCIAL TEÓRICO

A evasão pode ser definida como sendo o desligamento do aluno da instituição de ensino no qual o mesmo não retorna mais ao sistema escolar (SILVA FILHO e DE LIMA ARAÚJO, 2017). Segundo VIEIRA e GALINDO e CRUZ (2017), a evasão pode ter motivações de cunho pessoal (situação própria do estudante) ou institucional, sendo esta última dividida entre fatores internos ou externos à instituição. A situação sócio-econômica do aluno (e.g., baixa renda) é um exemplo de fator pessoal. Um fator

institucional interno seria a desmotivação do aluno por falta de estrutura (e.g, defasagem do acervo bibliográfico e equipamento laboratorial) e, um institucional externo, a ausência de transporte público para deslocamento à instituição.

Em 2008 foi sancionada a Lei nº 11.892/08, criando 38 Institutos Federais de Educação, Ciência e Tecnologia (IFET). Esta lei, juntamente com o cenário político à época contribuiu para a intensificação do processo de expansão e interiorização dos institutos federais (BRASIL, 2014). A expansão foi benéfica pois possibilitou o acesso a educação superior, profissional e tecnológica nos mais diversos níveis a uma significativa parcela da população. Entretanto, tamanha expansão trouxe implicações, como o crescimento da evasão discente.

A variedade de fatores motivadores possíveis, aliada a subjetividade discente, faz com que a evasão seja um problema difícil de se prever e lidar. Neste contexto uma análise temporal dos dados pode auxiliar a gestão da instituição a entender como os números da evasão se distribuem e se comportam ao longo do tempo, traçando estimativas futuras para um planejamento estratégico.

No campo computacional, técnicas de inteligência artificial vem sendo utilizadas compreensão e predição do processo de evasão discente (MDUMA and KALEGELE and MACHUVE, 2019). No âmbito do Instituto Federal do Ceará (IFCE), o trabalho em HORTÊNCIO FILHO e SILVA e LEAL (2020) avalia a assertividade de modelos obtidos de algoritmos clássicos de *machine learning*, que se baseiam em dados reais de alunos evadidos e não-evadidos. A ideia foi obter um feedback de modelos de predição construídos a partir de algoritmos de classificação tendo como base os dados de evasão da instituição.

O assunto de séries e análises temporais vêm sendo explorado ao longo dos anos nas mais diversas áreas. Na esfera educacional, HAIYANG et al. (2018) utiliza um método de classificação baseado em séries temporais aplicada ao contexto dos MOOCs (*Massive Open Online Courses*).

Análises de séries temporais também podem ser utilizadas no contexto de questões sociais. O trabalho descrito em REDHA (2018) utiliza técnicas de algoritmos genéticos e ARIMA em conjunto com intenção de incrementar a acurácia dos resultados para predição da taxa de evasão de escolas de primeiro grau do Iraque, considerando dados do período entre 2007 e 2015. O estudo conclui que a evasão de meninas nas escolas analisadas tende a ser maior do que a dos meninos e sugere que medidas de

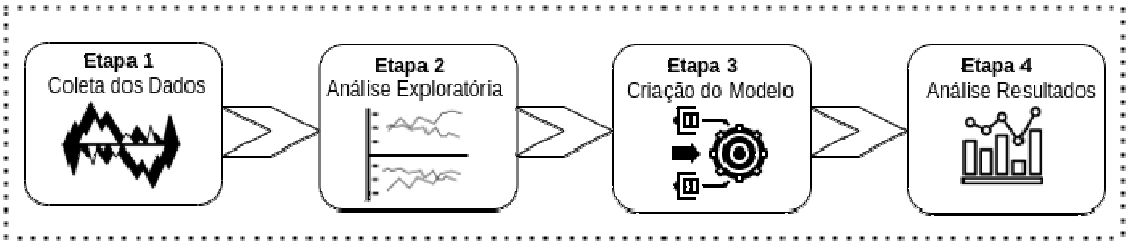
conscientização à população sejam adotadas em relação a importância da educação para mulheres.

O trabalho deste artigo tem como foco a análise de modelos temporais que visam predizer as taxas de evasão geral de instituições de ensino, não sendo objeto deste estudo métodos de classificação de evadidos. Para esse objetivo são utilizadas três técnicas de séries temporais baseadas em estratégias distintas, de forma a incrementar as chances de obtenção de bons resultados.

METODOLOGIA

Esta seção apresenta a metodologia aplicada ao longo do desenvolvimento deste trabalho que é resumida e ilustrada na Figura 1. Dentre os tópicos discutidos estão a forma como os dados foram obtidos e preparados, uma análise exploratória a partir destes dados,a definição dos modelos utilizados e a análise dos resultados obtidos.

Figura 1. Visão geral da metodologia utilizada



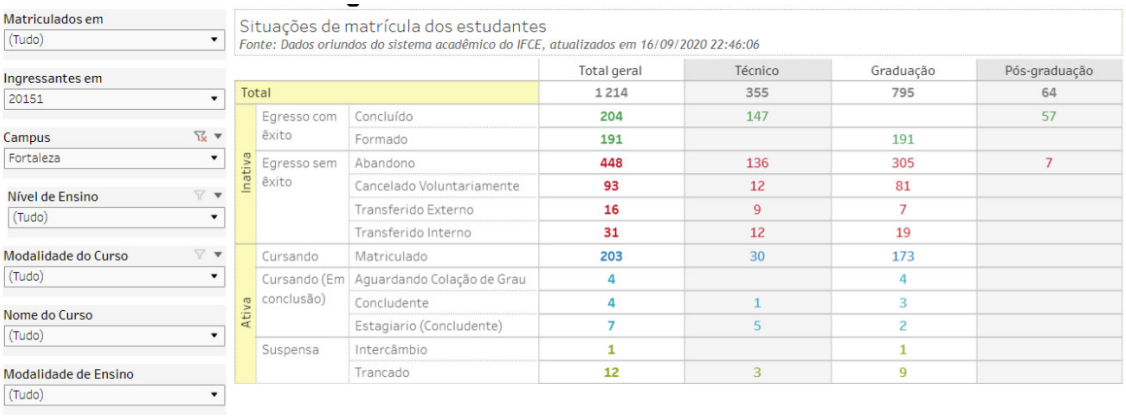
Etapa 1. O Instituto Federal do Ceará, utilizado como estudo de caso neste trabalho, criou a plataforma IFCE em Números (PRÓ-REITORIA DE ENSINO, 2017). Esta fornece uma visão sistêmica da instituição e seu contexto para auxiliar na elaboração e planejamento de ações, principalmente, no combate ao problema da evasão de estudantes. Para construção das séries temporais, foram utilizados dados diretamente extraídos do portal IFCE em números, com última atualização disponível em 16/09/2020 (Figura 2), relativos aos períodos semestrais de 2009.1 até 2017.2 dos nove campi de primeira geração (criados até 2008): Fortaleza, Iguatu, Crato, Juazeiro do Norte, Cedro, Maracanaú, Quixadá, Sobral e Limoeiro do Norte. Sendo m_i o total de alunos ingressantes no período i, e_i o número destes alunos que evadiram-se[4], a taxa de

[4] Considera-se evadido o aluno que abandonou, cancelou a matrícula, transferiu internamente (de um curso para outro dentro da própria instituição) ou externamente (para outra instituição).

evasão utilizada para a série temporal se dá por $T_x = (e_i/m_i)$ x 100. No exemplo da Figura 2, $e_i = 448 + 93 + 16 + 31$ e $m_i = 1214$, portanto $T_x = (588/1214)$ x 100 = 48,43%.

Etapa 2. Esta etapa consiste da análise exploratória das séries temporais. Para isso, foram concebidas as séries temporais respectivas a cada campus. A construção das séries, bem como demais implementações de técnicas, gráficos e validações, foi feita utilizando a linguagem R, juntamente com a ferramenta *Rstudio*[5]. Destaca-se o uso das bibliotecas *forecast, stats, seasonal e DescTools* para geração e avaliação dos modelos.

Figura 2. Plataforma IFCE em Números: dados do período 2015.1



Matriculados em: (Tudo)
Ingressantes em: 20151
Campus: Fortaleza
Nível de Ensino: (Tudo)
Modalidade do Curso: (Tudo)
Nome do Curso: (Tudo)
Modalidade de Ensino: (Tudo)

Situações de matrícula dos estudantes
Fonte: Dados oriundos do sistema acadêmico do IFCE, atualizados em 16/09/2020 22:46:06

			Total geral	Técnico	Graduação	Pós-graduação
Total			1 214	355	795	64
Inativa	Egresso com êxito	Concluído	204	147		57
		Formado	191		191	
	Egresso sem êxito	Abandono	448	136	305	7
		Cancelado Voluntariamente	93	12	81	
		Transferido Externo	16	9	7	
		Transferido Interno	31	12	19	
Ativa	Cursando	Matriculado	203	30	173	
	Cursando (Em conclusão)	Aguardando Colação de Grau	4		4	
		Concludente	4	1	3	
		Estagiario (Concludente)	7	5	2	
	Suspensa	Intercâmbio	1		1	
		Trancado	12	3	9	

Etapa 3. Esta etapa foca na concepção dos modelos temporais a partir das técnicas ARIMA (*autoregressive integrated moving average*), regressão linear aplicada a séries temporais (TSLM), e redes neurais autoregressivas (NNAR), amplamente aceitas e utilizadas em problemas de séries temporais. Na técnica ARIMA, foi considerada a parametrização que fornecesse os melhores ajustes de modelo[6] para para cada campus. Enquanto os dados dos semestres de 2009.1 à 2017.2 foram usados para construção do modelo (etapa de treino), os dados relativos à 2018.1 e 2018.2 foram utilizados como alvos de uma previsão temporal (teste), baseada nos modelos construídos, de forma a averiguar se os valores preditos são próximos aos valores reais.

Etapa 4. Por fim, a quarta etapa trata da análise dos resultados. Para esta finalidade, métricas tradicionalmente citadas na literatura (BRUCE AND BRUCE, 2019) foram utilizadas. Dentre elas destacam-se o MAPE (*Mean Absolute Percentage Error)* e o RMSE (*Root Mean Squared Error*), cujas fórmulas estão definidas na Equação 1 e Equação 2 respectivamente, onde N representa o número de períodos de tempo

[5] https://www.rstudio.com/
[6] Função auto.arima (https://otexts.com/fpp2/arima-r.html).

(semestres), a_j representa o valor real no tempo j e p_j representa o valor previsto para o instante de tempo j.

$$MAPE = \frac{1}{N}\sum_{j=1}^{N}\left|\frac{a_j - p_j}{a_j}\right| \times 100 \quad (1) \qquad RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(a_j - p_j)^2} \quad (2)$$

Além destas métricas, serão aplicados aos resíduos do modelo ARIMA os testes de hipóteses de auto-correlação de Ljung-Box, e aos resíduos do modelo de regressão linear o de Breusch-Godfrey. Para ambos os tipos de modelo também será aplicado o teste de normalidade Shapiro-Wilk. Tais testes tem como objetivo a validação estatísticas dos modelos, de forma averiguar a confiança do ajuste dos modelos e se há presença de enviesamentos.

RESULTADOS E DISCUSSÃO

A Tabela 1 com os dados da análise exploratória das taxas de evasão de cada campi. Os gráficos das séries temporais gerados para cada um dos 9 campi da 1ª geração do IFCE, após a extração dos dados no portal IFCE em números, são representados pela linha preta na Figura 3.

Tabela 1. Resultado das métricas de avaliação por campus e técnica utilizada

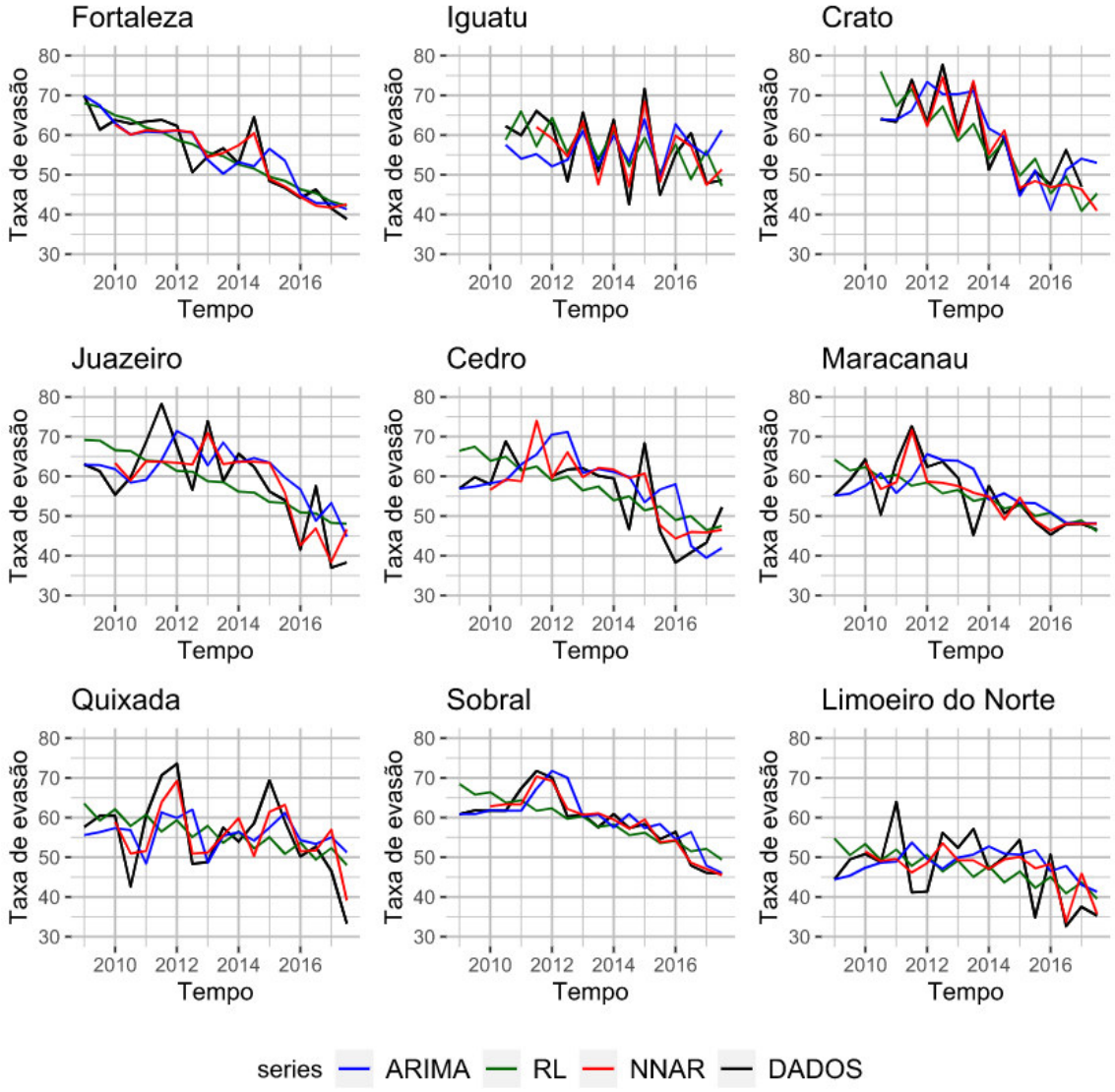
Evasão (em %)	For	Igu	Cra	Jua	Ced	Mar	Qui	Sob	Lim
Mínimo	38,86 (2017.2)	42,51 (2014.2)	45,49 (2015.1)	37,01 (2017.1)	38,30 (2016.1)	45,21 (2013.2)	**33,33 (2017.2)**	45,87 (2017.2)	32,66 (2016.2)
Média	55,16	58,03	**60,54**	58,63	56,95	55,18	55,78	58,96	47,12
Máximo	69,94 (2009.1)	71,63 (2015.1)	77,68 (2012.2)	78,22 (2011.2)	**80,77 (2011.2)**	72,57 (2011.2)	73,53 (2012.1)	71,75 (2011.2)	63,96 (2011.1)

A partir dos dados Tabela 1, constata-se que as taxas de evasão variaram, de maneira geral, entre 30% e 80%, chegando a 80.77% (campus Cedro em 2011.2). O menor valor percentual registrado foi no campus Quixadá (33.33% em 2017.2). A média geral é de 56.26%, que é considerada alta e reforça a importância do diagnóstico e o estabelecimento de ações para o combate a evasão.

Na Figura 3, é possível notar que as curvas têm uma leve tendência de queda, mais notórias nos campi de Fortaleza e Crato. Os gráficos resultantes dos modelos gerados através das técnicas ARIMA, Regressão linear para séries temporais (RL) e

redes neurais autoregressivas (NNAR), devidamente ajustados em relação aos dados reais das séries temporais são exibidos na Figura 3 para comparação. É possível perceber visualmente que, de maneira geral, os modelos conseguiram acompanhar as tendências das curva temporais reais (DADOS - linha preta).

Figura 3. Modelos ajustados às respectivas séries temporais.



A Tabela 2 exibe os resultados das métricas MAPE e RMSE para as curvas de cada gráfico, onde A é abreviação para ARIMA, RL para Regressão Linear e RN para Rede Neural. Os valores gerados a partir da métrica de MAPE foram todos abaixo de 15.42% (Campus Limoeiro), havendo casos com erro menor que 10% (resultados obtidos pela rede neural). Isto significa que, em média, os modelos tem um baixo percentual de erro. Com relação à métrica RMSE, os valores ajustados também foram

baixos, todos na casa de um dígito, tendo a rede neural alcançado resultados na ordem de 5% em sua maioria. É importante destacar as medidas RMSE baixas, pois representam o quão próximo os pontos dos valores previstos pelo modelo estão de dados reais.

Tabela 2. Resultado das métricas de avaliação por campus e técnica utilizada

Campus	RMSE-A	MAPE-A	RMSE-RL	MAPE-RL	RMSE-RN	MAPE-RN
Fortaleza	5.1960	7.1395	4.1445	5.2121	3.4212	4.6238
Iguatu	7.4178	12.2106	6.7858	10.4644	3.6065	6.1316
Crato	8.4948	12.9600	7.0937	11.0475	4.2681	6.4814
Juazeiro	8.9534	13.9771	8.4404	13.2379	6.1705	8.6087
Cedro	9.3076	13.1726	8.2672	12.1482	5.9565	8.4755
Maracanaú	6.6522	8.8864	5.8426	76076	4.2368	5.1031
Quixadá	8.8975	13.7108	8.8281	13.8136	5.6809	9.6146
Sobral	3.9242	4.8262	4.4141	5.9045	1.7310	2.2970
Limoeiro	8.3692	15.4253	7.2698	13.7706	5.7073	9.6610

A Tabela 3 exibe os valores resultantes dos testes de hipótese de autocorrelação Ljung-Box para resíduos dos modelos ARIMA, e Breusch-Godfrey para resíduos dos modelos de regressão linear, além do teste de normalidade Shapiro-Wilk, aplicados em ambos os casos.

Tabela 3. Testes de autocorrelação e normalidade nos resíduos

Teste	For	Igu	Cra	Jua	Ced	Mar	Qui	Sob	Lim
normalidade ARIMA	0.5407	0.3451	0.0733	0.6668	0.5803	0.7595	0.4664	0.3146	0.7812
normalidade RL	0.023	0.354	0.4902	0.1843	0.0848	0.2697	0.5135	0.5548	0.1768
Ljung-box	0.1658	0.4641	0.3877	0.8529	0.8385	0.4714	0.4921	0.226	0.6367
Breusch-Godfrey	0.67	0.05527	0.7828	0.7465	0.7701	0.3396	0.1154	0.1727	0.5024

No caso dos testes de autocorrelação, valores abaixo de 0.05 indicam que os dados residuais gerados são autocorrelacionados, tornando o modelo não confiável. Para o teste de normalidade, valores abaixo de 0.05 representam uma distribuição não-normal dos resíduos. A normalidade dos resíduos é considerada pré-requisito para que os resultados do ajuste dos modelo sejam confiáveis. Conforme disposto na Tabela 3, apenas o modelo de regressão linear do campus Fortaleza (valor sublinhado) obteve uma distribuição não-normal. Nos testes de autocorrelação, todos obtiveram valores acima de 0.05, i.e., os resíduos destes modelos não estão autocorrelacionados entre si.

Por fim, a Tabela 4 e Tabela 5 mostram resultados de uma previsão feita para os para os nove campi nos semestres 2018.1 e 2018.2 respectivamente, onde se destacam

em negrito os melhores valores para cada campus. Mais especificamente, essas tabelas mostram qual a previsão da taxa de evasão dentre os discentes que ingressaram em 2018.1 e 2018.2 para cada um dos nove campi. Os campi de Iguatu e Crato obtiveram os melhores valores preditos considerando a Regressão Linear e ARIMA respectivamente. Quixadá obteve a previsão mais distante do resultado real. Constata-se que os resultados previstos foram maiores do que os valores reais do momento em sua maioria (a exceção do campus Crato).

Tabela 4. Resultados preditos x valores reais para o período 2018.1

		For	Igu	Cra	Jua	Ced	Mar	Qui	Sob	Lim
Valores reais 2018.1		23.08	46.47	23.40	26.11	28.84	38.86	22.53	34.50	38.72
ARIMA 2018.1		**37.67**	60.87	**23.24**	**41.46**	47.41	47.37	46.62	**45.87**	**39.43**
RL 2018.1		40.16	**54.34**	36.56	45.67	**43.95**	**46.97**	50.88	50.06	42.27
RN 2018.1		39.58	82.86	42.67	45.48	48.14	48.11	**39.13**	49.38	45.51
95% (IC)	sup	26.52	34.98	5.36	22.84	21.78	31.30	27.20	37.95	38.22
	inf	48.83	73.70	41.13	60.07	66.13	62.64	74.55	53.78	61.90

Tabela 5. Resultados preditos x valores reais para o período 2018.2

		For	Igu	Cra	Jua	Ced	Mar	Qui	Sob	Lim
Valores reais 2018.2		24.11	40.28	13.42	35.62	25.22	29.87	25.19	30.56	19.30
ARIMA 2018.2		**35.76**	54.73	**11.26**	**41.46**	49.98	47.37	55.36	**45.87**	39.43
RL 2018.2		39.14	**45.53**	40.83	45.46	**44.95**	**44.25**	**46.64**	47.43	**38.15**
RN 2018.2		36.65	51.33	46.32	46.72	52.79	48.11	60.10	54.09	45.09
95% (IC)	sup	23.15	25.87	-14.02	20.49	22.78	28.58	22.96	34.67	35.59
	inf	48.37	65.19	36.55	62.42	67.12	59.92	70.32	57.06	59.27

Na prática, é extremamente difícil acertar uma previsão de maneira precisa. Intervalos de confiança são frequentemente calculados considerando os pontos de previsão como sendo o centro do intervalo. Assim, foram gerados, através das ferramentas estatísticas em linguagem R, os intervalos de confiança (IC) a 95% considerando os pontos de previsões que mais se aproximaram dos dados reais. Isto significa que há 95% de chances do valor real estar entre os valores de intervalo inferior e superior de confiança delimitado. Das 18 melhores previsões obtidas (em negrito), 5 ficaram fora do intervalo de confiança (Fortaleza/2018.1; Quixadá/2018.1; Sobral/2018.1; Sobral/2018.2 e Limoeiro do Norte/2018.2) e as outras 13 ficaram dentro do intervalo de confiança, o que representa 72,2% das previsões. Destas 13 previsões, 3 foram obtidas pelo ARIMA e 10 pela regressão linear.

A rede neural, apesar de obter os melhores índices de RMSE e MAPE de ajuste de modelo, não obteve as melhores previsões. Uma possível causa seria o fenômeno de *overfiting* (sobreajuste) da rede, que é quando os dados de treinamento não estão sendo suficientes para generalizar o aprendizado da rede ao que acontece no mundo real, o que poderia indicar a necessidade de mais dados de treinamento para a construção dos modelos.

Ressalta-se, ainda, que o valor real de evasão do ano de 2018 ainda pode sofrer um aumento, uma vez que alguns alunos ingressantes daquele ano ainda estão em curso podem vir a evadir. Isto pode aproximar ainda mais os valores reais das previsões, pois os pontos de previsão foram, predominantemente, maiores que os dados reais.

CONSIDERAÇÕES FINAIS

Este trabalho se propôs a avaliar o potencial de técnicas para construção de séries e modelos temporais no auxílio à previsão quantitativa de evasões discentes. Os dados foram coletados do portal IFCE em números e dispostos em uma frequência semestral. Foram utilizadas três técnicas presentes na literatura (Regressão linear, ARIMA e redes neurais NNAR) para a geração dos modelos de previsão. Para avaliação dos valores ajustados, foram utilizadas as métricas RMSE e MAPE.

Os resultados demonstraram que a maioria dos valores reais foram contemplados nos intervalos de confiança obtidos a partir dos melhores pontos de previsão, indicando bom potencial no uso destas técnicas para situações reais.

Como trabalhos futuros, pretende-se obter dados mais antigos (2008 e anteriores) de forma a incrementar os resultados das séries temporais, em especial, os obtidos pela rede neural. Também pretende-se utilizar outras técnicas para construção e avaliação de novos modelos, como ETS, Holt/Holt-Winter, bem como explorar outros campi de gerações seguintes, como o campus Aracati, Baturité e Canindé.

Por fim, acredita-se que é viável o desenvolvimento de uma ferramenta para exibição das informações oriundas de uma análise exploratória para cada campus, bem o cálculo, predição dos níveis de evasão por campus no qual possa ser utilizada por gestores para fins de diagnóstico e previsão dos níveis de evasão discente.

AGRADECIMENTOS

Ao Instituto Federal do Ceará por todo o apoio no desenvolvimento dessa pesquisa, em especial, ao diretor geral do Campus Paracuru, Toivi Masih Neto, por todo o apoio e incentivo prestado.

REFERÊNCIAS

ABDEL-AAL, R. E.; MANGOUD, A. M. Modeling and forecasting monthly patient volume at a primary health care clinic using univariate time-series analysis. Computer methods and programs in biomedicine, v. 56, n. 3, p. 235-247, 1998.

ARIYO, Adebiyi A.; ADEWUMI, Adewumi O.; AYO, Charles K. Stock price prediction using the ARIMA model. In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014. p. 106-112.

BOX, George EP et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

BRASIL (2014). Documento orientador para a superação da evasão e retenção na Rede Federal de Educação Profissional, Científica e Tecnológica. Ministério da Educação. Secretaria de Educação Profissional e Tecnológica.

BRUCE, Andrew; BRUCE, Peter. Estatística Prática para Cientistas de Dados. Alta Books, 2019.

HORTÊNCIO FILHO, Fernando Wagner Brito; VINUTO, Tiago Silva; LEAL, Bruno Carvalho. Análise de Classificadores para Predição de Evasão dos Campi de uma Instituição de Ensino Federal. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. SBC, 2020. p. 1132-1141.

HAIYANG, Liu et al. A time series classification method for behaviour-based dropout prediction. In: 2018 IEEE 18th international conference on advanced learning technologies (ICALT). IEEE, 2018. p. 191-195.

VIEIRA, Armênia Chaves Fernandes; GALLINDO, E. de L.; CRUZ, H. A. Plano estratégico para permanência e êxito dos estudantes do IFCE. Fortaleza: IFCE, 2017.

JIANG, Shancheng et al. Combining Deep Neural Networks and classical time series regression models for forecasting patient flows in Hong Kong. IEEE Access, v. 7, p. 118965-118974, 2019.

MDUMA, Neema; KALEGELE, Khamisi; MACHUVE, Dina. A survey of machine learning approaches and techniques for student dropout prediction. 2019.

Pró-reitoria de ensino do IFCE (2017). IFCE em números. Disponível em: http://ifceemnumeros.ifce.edu.br/. Acesso em 25 de jun de. 2021.

REDHA, Sabah Manfi. The Prediction of the Rate of the Dropout of the Primary Schools Students by Using the Genetic Algorithm. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, v. 9, n. 2, p. 198-214, 2018.

SANTOS, Pedro Vieira Souza. PREVISÃO DA DEMANDA POR PRODUÇÃO DE CAFÉ NO BRASIL: UMA ANÁLISE. Latin American Journal of Business Management, v. 11, n. 1, 2020.

SILVA FILHO, Raimundo Barbosa; DE LIMA ARAÚJO, Ronaldo Marcos. Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. Educação por escrito, v. 8, n. 1, p. 35-48, 2017.