

V SEMANA UNIVERSITÁRIA DA URCA

XXIII Semana de Iniciação Científica

07 a 11 de Dezembro de 2020

Tema: “Os impactos e desafios da pandemia COVID no ensino, pesquisa e extensão”



PREVISÃO DO RISCO DE REPROVAÇÃO NO ENSINO SUPERIOR: UMA APLICAÇÃO DOS ALGORITMOS DE MACHINE LEARNING PARA A UNIVERSIDADE REGIONAL DO CARIRI

Pedro Willame Pereira da Silva¹, Jayane Freires Ferreira², José Alexandre Queiroga do Nascimento³ Andréa Ferreira da Silva⁴

Resumo: A reprovação no ensino superior tem se caracterizado como um problema nas Instituições de Ensino Superior (IES) no Brasil, o aumento de ofertas de vagas na educação superior em decorrência dos investimentos em educação evidencia essa afirmativa. Nesse sentido este artigo tem como objetivo prever o risco de reprovação dos discentes da Universidade Regional do Cariri (URCA), Unidade Descentralizada de Iguatu. Utilizando-se microdados cedidos pela URCA, irão ser aplicados modelos e algoritmos de *Machine Learning* (ML), para a análise dos dados e estimação das previsões. Os resultados obtidos parcialmente evidenciam que entre as disciplinas que mais reprovam figuram as disciplinas pesquisa, e que pode haver uma possível relação entre as variáveis exploradas até o momento e o desempenho acadêmico dos discentes. A conclusão parcial é de que as possíveis correlações entre as variáveis de controle e o problema em questão, partir dos modelos de previsões aplicados posteriormente, podem auxiliar em medidas de caráter preventivo no auxílio das políticas de intervenção.

Palavras-chave: Machine Learning. Economia. Políticas Públicas. Reprovação.

1. Introdução

Problemas como a reprovação, a retenção e a evasão incidem sobre uma realidade presente na formação de muitos discentes nos mais diversos cursos de formação nas Instituições de Ensino Superior (IES) no Brasil. Caracterizam-se não só como dificuldades estruturais do contexto educacional, mas acarretam também um aumento nos custos de financiamento do funcionamento nas universidades públicas, diminuindo os níveis de produtividade e eficiência destas. Diversas pesquisas abordam essa problemática, dialogando sobre a necessidade de interferência em diferentes níveis, partindo desde o núcleo familiar, social e econômico, até ao que remete aos insumos escolares e diretrizes da educação (BARROSO; FALCÃO, 2004; GAIOSO, 2005; GISI, 2006; SILVA-FILHO et al., 2007; SANTOS-BAGGI; LOPES, 2011).

Nessa perspectiva, a partir dos dados oriundos do Censo da Educação Superior referentes ao ano de 2018, divulgados pelo Instituto Nacional de

1 Graduando em Ciências Econômicas pela Universidade Regional do Cariri (URCA) – Unidade Descentralizada do Iguatu (UDI). e-mail: pedrowill2@hotmail.com

2 Graduanda em Ciências Econômicas pela Universidade Regional do Cariri (URCA) – Unidade Descentralizada do Iguatu (UDI). e-mail: jayfreires2014@gmail.com

3 Graduando em Ciências Econômicas pela Universidade Regional do Cariri (URCA) – Unidade Descentralizada do Iguatu (UDI). e-mail: jose.queirogan@gmail.com

4 Doutora em Economia Aplicada pela Universidade Federal da Paraíba (UFPB). Docente do Curso de Ciências Econômicas da Universidade Regional do Cariri, e-mail: andrea.economia@yahoo.com.br

V SEMANA UNIVERSITÁRIA DA URCA

XXIII Semana de Iniciação Científica

07 a 11 de Dezembro de 2020

Tema: "Os impactos e desafios da pandemia COVID no ensino, pesquisa e extensão"



Estudos e Pesquisa Anísio Teixeira (INEP), no Brasil, há 2.537 IES, com 37.962 cursos de graduação ofertando 13,5 milhões de vagas, com relação ao censo de 2017 houve um aumento de cerca de 20,7% no número de vagas ofertadas. Mesmo nesse cenário que enfatiza a expansão da educação superior, enquanto o número de concluintes em cursos de graduação no ensino superior no ano de 2017 foi de aproximadamente 1,2 milhões, em 2018 esse número foi de apenas cerca de 1,26 milhões de alunos, representando uma variação positiva de 5,2% na quantidade de discentes concluintes entre os anos citados.

Sendo a evasão caracterizada como um dos problemas que afligem as instituições em geral, estando correlacionada com os índices de reprovação que acabam por afetar os resultados dos sistemas educacionais, as consequências dessas recorrentes evasões acarretam em desperdícios sociais, acadêmicos e econômicos. No setor público são recursos investidos sem um retorno satisfatório que culminam na diminuição da taxa de retorno de investimentos em educação, no setor privado além da perda de receitas, causa um deslocamento no que poderia ser aproveitado como mão de obra qualificada, através de profissionais formados e inseridos no mercado de trabalho (SILVA-FILHO et al., 2007).

Tendo em vista o contexto abordado, a reprovação não deve ser analisada de uma forma superficial, e sim com a identificação de fatores que possam estar influenciando o aluno até esse quadro de baixo desempenho. É importante observar tais fatores tanto na ótica do aluno quanto da instituição, assim, enfatizando a necessidade de elaborar medidas de intervenção que visem diminuir retenções, reprovações e evasões, e consequentemente custos gerados por tais problemas. Nesse contexto, a análise de previsão se torna uma ferramenta importante no que remete a elaboração de políticas públicas voltadas a resolução dessas questões. Mas como executar a previsão do risco de reprovação, quais são os principais fatores e como determinar medidas preventivas diante desse cenário?

2. Objetivo

Prever os riscos de reprovação dos discentes a partir das disciplinas com os maiores índices de retenções nos quatro cursos que compõem a Universidade Regional do Cariri (URCA), Unidade Descentralizada de Iguatu (UDI). Descrevendo os principais algoritmos de *Machine Learning* (ML) que serão adotados para executar as previsões e classificando os discentes com maior risco de reprovação a partir da análise dos algoritmos aplicados.

3. Metodologia

A metodologia que será adotada para a análise dos dados presentes na pesquisa serão as aplicações de técnicas de *Machine Learning* (ML), tendo em vista a necessidade da análise de uma base de dados, acadêmicos e socioeconômicos, oriundos do sistema de informação da Universidade Regional do Cariri (URCA), Unidade Descentralizada de Iguatu (UDI). Foram coletados os

V SEMANA UNIVERSITÁRIA DA URCA

XXIII Semana de Iniciação Científica

07 a 11 de Dezembro de 2020

Tema: “Os impactos e desafios da pandemia COVID no ensino, pesquisa e extensão”



registros administrativos e acadêmicos da URCA/UDI referentes aos discentes matriculados em todas as disciplinas referentes aos quatro cursos que compõem esta IES, de 2010 a 2019.

Dada a complexibilidade de se analisar esse conjunto de variáveis coletadas até o presente momento da pesquisa, a opção pela aplicação dos modelos de ML consiste no interesse por modelos preditivos, empregados como ferramentas de compreensão e modelagem de dados complexos. Inicialmente foram realizadas análises com o auxílio do *software* R, linguagem utilizada na leitura e visualização de dados e estimações estatísticas. Onde foram examinados os registros acerca de reprovações por disciplinas e dados referentes aos desempenhos de discentes, com a finalidade de verificar a possibilidade de existência de correlações iniciais referentes ao objetivo proposto da pesquisa.

4. Resultados preliminares

Apresentando os dados parciais uma vez que os registros coletados provisoriamente junto ao sistema de informação da URCA/UDI, mostraram-se carentes de variáveis que pudessem corroborar o objetivo central desta pesquisa, as demonstrações a seguir se pautam dentro das análises metodológicas iniciais aplicadas com os dados disponíveis e suas inferências com a ideia central do artigo.

A Tabela 6 apresenta evidências sobre a variável resposta, reprovado ou aprovado, e das variáveis que representam o desempenho do aluno, média do vestibular em redação, quantidade de alunos e sua classificação, curso de graduação e média final nas disciplinas de pesquisa⁵ que mais reprovam nos quatro cursos que compõem a URCA/UDI. Assim, estão expostos na Tabela 6 os testes de média e intervalo de confiança das variáveis citadas anteriormente, correspondente aos dados de registros de 2010 a 2019.

Um dos pressupostos adotados, tomando como base a visualização dos resultados parciais, se baseia na hipótese que os estudantes que obtiveram uma menor média na disciplina de Redação no vestibular, conseqüentemente figuram entre os alunos reprovados. Isso evidencia uma possível relação entre as variáveis, onde o resultado em uma disciplina do vestibular infere em um déficit no resultado em uma disciplina específica na graduação.

Considerando as taxas de reprovação, Sampaio *et al.* (2011) discorre sobre um ponto correlato ao objetivo desta pesquisa, onde cita que a evasão incide além das vagas ociosas nas perdas em relação aos investimentos educacionais gerando um aumento nos custos. Os resultados parciais demonstram a necessidade do aprofundamento da análise dos dados em função de um

⁵ As disciplinas por curso são: Ciências Econômicas: elaboração de monografia I; elaboração de monografia II; elaboração de monografia III; metodologia da pesquisa econômica; metodologia do trabalho científico; técnicas de pesquisa econômica; técnicas de pesquisa em economia. Direito: metodologia do trabalho científico; metodologia do trabalho científico; monografia I (projeto de monografia); monografia II (monografia). Educação física: monografia I; monografia II. Enfermagem - trabalho monográfico I (projeto); trabalho monográfico II (relatório); metodologia da pesquisa; metodologia do trabalho científico.

V SEMANA UNIVERSITÁRIA DA URCA

XXIII Semana de Iniciação Científica

07 a 11 de Dezembro de 2020

Tema: "Os impactos e desafios da pandemia COVID no ensino, pesquisa e extensão"



compreendimento sobre o problema da reprovação, baseando-se em uma estratégia empírica e embasado em uma estrutura metodológica concisa.

Tabela 6 – Evidências iniciais da variável de resultado, por status de matrícula, dos discentes da URCA\UDI, 2010 a 2019 – Teste de médias e intervalo de confiança¹.

Curso	Reprovado	Alunos	Média Idade	Média Vest. Redação	Média final disciplina
Ciências econômicas	Não	1071	27.47 (27.16 - 27.78)	5.459 (5.343 - 5.576)	8.706 (8.647 - 8.766)
	Sim	102	28.34 (26.95 - 29.72)	5.164 (4.766 - 5.562)	0.254 (0.099 - 0.410)
Direito	Não	1048	28.56 (28.11 - 29.00)	7.012 (6.909 - 7.116)	9.049 (8.994 - 9.104)
	Sim	47	30.34 (28.82 - 31.85)	6.547 (6.026 - 7.068)	0.446 (0.141 - 0.751)
Educação física	Não	444	27.64 (27.15 - 28.12)	5.377 (5.202 - 5.553)	8.826 (8.744 - 8.908)
	Sim	58	32.03 (30.12 - 33.94)	5.241 (4.735 - 5.747)	0.310 (0.068 - 0.552)
Enfermagem	Não	1088	26.12 (25.86 - 26.38)	5.922 (5.810 - 6.034)	8.986 (8.928 - 9.043)
	Sim	15	27.40 (25.11 - 29.68)	5.266 (4.270 - 6.262)	0.800 (0.131 - 1.468)

Fonte: Elaboração própria a partir do microdados do Departamento de Tecnologia da Informação (DTI) da Universidade Regional do Cariri (URCA).

Notas: (1) Intervalos com 95% de confiança para as médias calculadas. Todos os intervalos de confiança apresentam um teste t de 0, 000.

5. Conclusão

A reprovação se demonstra como um problema da conjectura educacional, que afeta tanto o discente quanto a estrutura de investimentos e produtividade do capital envolvido no fomento da educação superior, além disso tal questão incide diretamente nos números relativos as taxas de evasão no ensino superior. Os resultados parciais da pesquisa mostraram inicialmente uma possível coesão entre as ideias de correlação presentes na metodologia, uma vez que se pode notar entre as disciplinas com o maior número de reprovações, uma grande parte está ligada as disciplinas ligadas na área de pesquisa.

Dessa maneira, conclui-se, parcialmente, que pode haver correlações existentes referentes as taxas de reprovações mediante a análise de uma série de variáveis a serem coletadas posteriormente com acesso ao banco de dados necessário. Assim, a previsão da reprovação através das ferramentas de ML pode se qualificar como um método viável a elaboração de políticas públicas educacionais e apoios aos gestores responsáveis pelas aplicações de tais políticas.

V SEMANA UNIVERSITÁRIA DA URCA

XXIII Semana de Iniciação Científica

07 a 11 de Dezembro de 2020

Tema: "Os impactos e desafios da pandemia COVID no ensino, pesquisa e extensão"



6. Agradecimentos

Agradeço a(o) orientador(a) Profa. Dra. Andréa Ferreira da Silva, pela orientação e condução das atividades de forma excepcional, aos colegas de curso e pesquisa, Jayane Freires Ferreira e José Alexandre Queiroga do Nascimento, pelo companheirismo e colaboração no desenvolvimento das atividades e a Universidade Regional do Cariri pela oportunidade de execução da pesquisa científica.

7. Referências

BARROSO, Marta F.; FALCÃO, Eliane BM. Evasão universitária: o caso do Instituto de Física da UFRJ. **IX Encontro Nacional de Pesquisa em Ensino de Física**, v. 9, p. 1-14, 2004.

GAIOSO, Natália Pacheco de Lacerda. **O fenômeno da evasão escolar na educação superior no Brasil**. 2005. 75 f. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, 2005.

GISI, Maria Lourdes. A Educação Superior no Brasil e o caráter de desigualdade do acesso e da permanência. **Revista Diálogo Educacional**, v. 6, n. 17, p. 97-112, 2006.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Sinopse Estatística da Educação Superior 2018. Brasília: Inep, 2019. Disponível em: <<http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse>>. Acesso em: 30.09.2020.

SAMPAIO, B.; SAMPAIO, Y.; MELLO, E. P. G. de; MELO, A. S. **Desempenho no vestibular, background familiar e evasão: evidências da UFPE**. **Economia Aplicada**, v. 15, n. 2, p. 287-309, 2011.

SANTOS BAGGI, Cristiane Aparecida Dos; LOPES, Doraci Alves. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 16, n. 2, p. 355-374, 2011.

SILVA FILHO, Roberto Leal Lobo et al. A evasão no ensino superior brasileiro. **Cadernos de pesquisa**, v. 37, n. 132, p. 641-659, 2007.