

Aplicação de Tecnologia de Identificação Preventiva de Risco de Evasão Escolar: Comparativo de Resultados em Níveis de Ensino Distintos

Leonardo da Silva Souza, Dr.

UFSB

Resumo

Este artigo apresenta uma análise dos relatórios estatísticos sobre a aplicação da tecnologia SACA+ para identificar preventivamente o risco de evasão escolar. O primeiro relatório, de 2016, descreve a aplicação da tecnologia em dados de frequência e pontualidade de 100 estudantes do Ensino Médio ao longo de 18 meses, concluindo que a ferramenta identificou com 75% de precisão os estudantes em risco de abandono no universo monitorizado. Posteriormente a tecnologia foi aprimorada e seus resultados foram relatados em um segundo relatório, focando num estudo realizado no Ensino Fundamental II da Escola Municipal Professor Álvaro Henrique entre agosto e dezembro de 2022, onde a tecnologia foi aplicada à frequência de 245 estudantes e previu corretamente 94% dos casos de evasão real. Ambos os estudos fornecem dados sobre a eficácia do SACA+ na previsão da evasão escolar, embora com diferentes percentagens de sucesso e contextos de aplicação.

Introdução

A evasão escolar representa um desafio significativo para os sistemas educacionais, impactando negativamente o desenvolvimento individual dos estudantes e o progresso social. A identificação preventiva de estudantes em risco de abandono escolar é crucial para permitir a implementação de ações preventivas e de apoio. Neste contexto, tecnologias baseadas em análise de dados de frequência têm sido propostas como ferramentas potenciais para auxiliar escolas na identificação preventiva desses estudantes.

O presente artigo tem como objetivo correlacionar e analisar os resultados de duas aplicações de uma tecnologia de identificação preventiva de risco de evasão escolar (SACA+ em um estudo e tecnologia desenvolvida pelo mesmo autor em outro estudo) em diferentes contextos educacionais: no Ensino Fundamental II de uma escola municipal e no Ensino Médio, integrado a um curso técnico. Busca-se, assim, compreender a performance da tecnologia e os fatores associados à sua aplicação em distintas realidades escolares.

Com base nas informações fornecidas nas fontes:

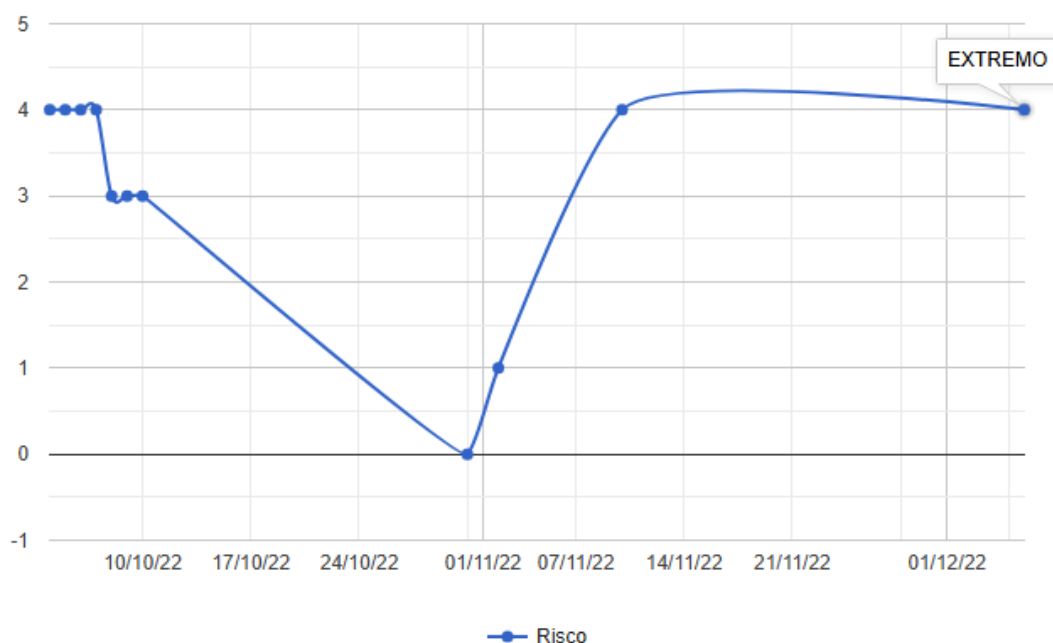
No primeiro relatório, "Estudo estatístico prévio base 2019", a tecnologia de identificação preventiva de risco de evasão escolar foi aplicada sobre dados de 100 estudantes de Ensino Médio público. Dentre os estudantes monitorados após a etapa inicial, 63 estudantes foram acompanhados nas turmas FT, WD e VD para a análise final. Para otimizar a aplicação da tecnologia, foram utilizadas amostragens variadas, incluindo análises com 99 estudantes, 21 estudantes, 18 estudantes, e 22 estudantes, além de uma avaliação de 560

percursos de teste. No segundo relatório, "Estudo Estatístico Colégio Municipal Álvaro Henrique", foram analisados dados de 245 estudantes pela tecnologia SACA+ na Escola Municipal Professor Álvaro Henrique. Este foi o universo de estudantes acompanhados pelo sistema no Ensino Fundamental II.

Para cada um destes estudantes, foram coletados dados de frequência e pontualidade diária nas aulas por intermédio dos professores de cada turma. A frequência se refere à presença do estudante em um dia, auferida por um professor. O atraso se refere à chegada do estudante após o período de tolerância de 15 minutos. Ao final, são utilizados uma sequência de marcadores de presença, tais como uma sequência de caracteres que indica presença, ausência, atraso ou falta na aula. Foi realizada a aplicação do algoritmo de classificação de risco sobre os dados de entrada para cada estudante monitorado ao final de cada semana. As classificações de risco são emitidas pela tecnologia em 5 níveis de precisão sendo eles: BAIXO, MODERADO, CONSIDERAVEL, ELEVADO, EXTREMO.

Com esta classificação semanal foram gerados gráficos que indicam a progressão do risco no tempo, tal como ilustrado na imagem 1, apresentando a análise de risco do SACA+ para cada estudante evadido ao longo das semanas de monitoramento.

Imagem 1: Exemplo de gráfico gerado pela tecnologia de identificação de risco de evasão



O risco se inicia em extremo e ao longo da linha do tempo ele reduz no dia 01/11/2022 para risco baixo, mas logo em seguida vai subindo até chegar ao risco extremo entre os dias 07/11 e 14/11/2022 e se mantém até o final do período.

Os principais desafios pontuados ou implícitos nos relatórios em questão incluem limitações nos dados disponíveis para análise, o impacto das ações de intervenção da escola na redução da evasão prevista pela tecnologia (o que também explica a diferença entre o número de estudantes em risco identificado e o número de evadidos reais), a gestão de um

número elevado de alertas de risco (falsos positivos), e a necessidade de otimizar a configuração da tecnologia para obter os melhores resultados.

Resultados Encontrados

Os resultados de ambos os estudos demonstram a capacidade da tecnologia em identificar estudantes em risco de evasão.

No estudo prévio de base 2019 (Ensino Médio):

- A tecnologia identificou o risco de evasão com uma precisão de 75% dentre o grupo de estudantes evadidos (conforme o resumo).
- 80% dos evadidos monitorados foram identificados com risco no período inicial (primeiros 6 meses), e entre 30% e 44% foram identificados no período final nas diferentes turmas.
- A classificação de risco após 18 meses indicou que 34% dos estudantes monitorados (63 no total) apresentavam risco EXTREMO de evasão. A evasão real no grupo total inicial foi de 27%.
- Uma parte significativa dos estudantes indicados em risco pela tecnologia não evadiu (56% na turma VD, 70% na FT, 68% na WD).
- Configurações otimizadas da tecnologia demonstraram maior performance: a melhor configuração apresentou Sensibilidade de 100% (identificando todos os evadidos reais na amostra de teste otimizada) e Precisão de 83,33%.

No estudo da Escola Municipal Professor Álvaro Henrique (Ensino Fundamental II):

- 94% dos estudantes que de fato evadiram foram previamente identificados pelo SACA+ como estando em risco extremo de evasão.
- Estes estudantes foram identificados com uma antecedência de 4 meses.
- 100% dos evadidos foram identificados com antecedência, sendo que os primeiros sinais de risco surgiram nas primeiras semanas de monitoramento.
- A tecnologia SACA+ apresentou uma margem de acerto de 87,5% em relação ao risco extremo de evasão nos estudantes evadidos.
- A distribuição de risco ao final do período indicou que 64,3% dos 245 estudantes se encontravam em risco extremo. A evasão real neste grupo foi de 6,5%.
- Dos 42 estudantes em recuperação, 28 (66,66%) foram classificados pelo SACA+ em risco extremo.
- A evasão foi maior entre estudantes do sexo feminino (62,5% dos evadidos) do que masculino (37,5%), desproporcionalmente à sua representação na população estudada (51,8% feminino, 48,2% masculino).

Análise dos Dados

O primeiro estudo, intitulado "Relatório estatístico da aplicação da tecnologia SACA+ na Escola Municipal Professor Álvaro Henrique", descreve a aplicação da tecnologia SACA+ na Escola Municipal Professor Álvaro Henrique, em Porto Seguro, focada no Ensino Fundamental II. O período de acompanhamento prático ocorreu de agosto a dezembro de

2022, dentro de um período de trabalho mais amplo (Novembro de 2021 a Dezembro de 2022). Foram analisados dados de frequência escolar de 245 estudantes. A entrada dos dados de presença ou ausência em sala de aula foi realizada pelos professores. O sistema SACA+ processou esses dados para indicar a progressão ou retração do risco de evasão de cada estudante, além de agregados de risco por turma. A disciplina específica na qual o sistema foi testado foi Artes. A população estudada neste relatório consistia em 51,8% mulheres e 48,2% homens.

O segundo estudo, intitulado "Estudo estatístico prévio base 2019", apresenta uma síntese da aplicação de uma tecnologia similar, elaborada pelo mesmo autor, no Ensino Médio público. Foram utilizados dados de frequência e pontualidade diária coletados por professores para 100 estudantes participantes de um curso técnico de nível médio integrado ao Ensino Médio, durante 18 meses. Os dados de entrada incluíam marcadores para presença, ausência, atraso (considerado uma forma de presença para outro algoritmo) ou falta. Inicialmente, havia 100 estudantes em uma turma FG na primeira etapa de 6 meses, após a qual os estudantes se dividiram em turmas WD, FT, VD e DG, sendo que os dados da turma DG não foram monitorados pela tecnologia.

Ambos os estudos utilizaram dados de frequência como base para a análise de risco, com entrada de dados pelos professores. No entanto, diferem no nível de ensino (Fundamental II vs. Médio), no tamanho da amostra (245 vs. 100 estudantes), no período de acompanhamento (meses vs. 18 meses), e na granularidade dos dados (presença/ausência vs. presença, ausência, atraso, falta). A classificação de risco no estudo de Ensino Médio utilizava 5 níveis (BAIXO, MODERADO, CONSIDERAVEL, ELEVADO, EXTREMO), enquanto o estudo de Ensino Fundamental II parece focar mais nos níveis baixo e extremo em alguns exemplos, mas apresenta a distribuição em 4 níveis (BAIXO, MODERADO, ELEVADO, EXTREMO) na análise final.

A evasão real reportada pela escola no estudo do Ensino Fundamental II foi de 17 estudantes (6,9%) no universo dos 245 acompanhados, sendo 1 transferido, totalizando 16 evadidos. No estudo do Ensino Médio, a evasão total informada pela escola foi de 37 estudantes no universo inicial de 100, sendo que 21 destes evadidos estavam no grupo monitorado pela tecnologia. A taxa de evasão real sobre o total de estudantes no estudo de Ensino Médio foi de 27,0%.

Correlação dos Resultados

Ambos os estudos indicam que a tecnologia é eficaz na identificação antecipada de estudantes que evadirão. O estudo mais recente (Álvaro Henrique) reporta que 94% dos evadidos foram identificados com risco extremo, enquanto o estudo anterior (Ensino Médio) menciona uma precisão de 75% entre evadidos e demonstra a possibilidade de alcançar 100% de sensibilidade e 83,33% de precisão com otimização de parâmetros. A capacidade de antecipação é confirmada em ambos, com 4 meses no EFII e até 12 meses no EM.

Um ponto comum e relevante é a constatação de que a tecnologia identifica um número consideravelmente maior de estudantes em alto risco do que o número real de evadidos. No estudo do EFII, 64,3% foram classificados em risco extremo, enquanto apenas 6,5% evadiram. No estudo do EM, 34% foram classificados em risco extremo (no grupo monitorado), enquanto a evasão total foi de 27% e a evasão no grupo monitorado foi de 21

evadidos em 63 monitorados (~33%). Esta diferença sugere que as ações preventivas e de suporte implementadas pelas escolas podem estar desempenhando um papel fundamental na retenção dos estudantes identificados em risco. No estudo do EFII, o fato de 66,66% dos estudantes em recuperação terem sido classificados em risco extremo reforça essa hipótese, indicando que a tecnologia identifica um grupo de estudantes que, de outra forma, poderiam abandonar, mas que são retidos por outras ações da escola. O estudo do EM também menciona a ocorrência de ações de orientação e busca ativa que podem ter impactado os resultados.

A análise por gênero no estudo do EFII apontou uma maior taxa de evasão entre meninas do que sua representação na população geral da escola, um fator demográfico não explorado nos resultados do estudo do EM nas fontes fornecidas.

Trabalhos Futuros

Com base nas informações contidas nas fontes, é possível identificar alguns desafios ou fatores que influenciaram os resultados da aplicação da tecnologia de identificação preventiva de risco de evasão escolar:

1. **Limitações na Coleta e Disponibilidade de Dados:** Na segunda fonte, é explicitamente mencionado que os dados de uma das turmas (DG), composta por 38 estudantes, não foram fornecidos, impossibilitando o acompanhamento da aplicação da tecnologia neste grupo. Além disso, devido à não adesão de alguns professores, nem todos os estudantes que evadiram foram de fato monitorados pela tecnologia (apenas 69% dos evadidos o foram neste estudo). Estas limitações impactam o universo total de estudantes analisados no estudo e a abrangência do monitoramento, tornando a análise menos completa do que seria ideal. Embora a primeira fonte mencione a entrada de dados por professores, não especifica se houve desafios significativos com a adesão ou completude dos dados neste caso específico.
2. **Impacto das Intervenções da Escola:** Ambas as fontes sugerem que as ações implementadas pelas escolas para reduzir o abandono e auxiliar os estudantes em risco tiveram um papel significativo nos resultados observados. Na primeira fonte, indica-se que, enquanto o SACA+ identificou 64,3% dos estudantes em risco extremo em dezembro de 2022, a evasão real foi de apenas 6,5%. A fonte atribui essa diferença, em parte, ao facto de as ações da escola terem auxiliado na manutenção da taxa de matrícula, mantendo grande parte dos estudantes identificados em risco. Similarmente, a segunda fonte nota que ações de orientação de percurso e busca ativa foram realizadas durante o período de aplicação da tecnologia e que isso "podendo ter impacto nestes números", referindo-se ao elevado número de estudantes identificados em risco pela tecnologia que não evadiram (entre 56% e 70% nas turmas monitoradas). Este fator torna a interpretação direta da precisão da tecnologia como preditora de *evasão real* mais complexa, pois a tecnologia identifica um risco que, idealmente, deve levar à intervenção, e o sucesso dessa intervenção reduz a evasão. A segunda fonte sugere mesmo que estudos futuros sem atividades de redução da evasão seriam úteis para

verificar o impacto da tecnologia minimizando sua margem de erro.

3. Elevada Taxa de Alerta (Falsos Positivos): Relacionado ao ponto anterior, os resultados em ambas as fontes mostram que a tecnologia tende a identificar um número consideravelmente maior de estudantes em risco do que o número que, de facto, evade. Embora isso possa ser influenciado pelas intervenções, também representa um desafio para a escola, que precisa gerir e intervir junto a um grupo maior de estudantes do que aqueles que acabam por evadir. Por exemplo, na primeira fonte, 64,3% foram classificados em risco extremo, enquanto 6,5% evadiram. Na segunda fonte, 34% dos monitorados foram classificados em risco extremo no período final, enquanto a evasão total foi de 27% (sem especificar a evasão apenas no grupo monitorado), e uma alta percentagem de estudantes identificados em risco não evadiu.
4. Necessidade de Otimização e Configuração: A segunda fonte descreve o processo de otimização da aplicação da tecnologia, variando parâmetros como o número de aulas analisadas e o tamanho das amostragens. Os resultados apresentados mostram que a sensibilidade, seletividade, precisão e erro variam significativamente dependendo da configuração utilizada. Encontrar a configuração "ótima" para a aplicação da tecnologia representa um desafio técnico para garantir que os resultados sejam os mais eficazes possíveis na identificação do risco.

Com base nos resultados e correlações, sugere-se que trabalhos futuros possam focar em:

- Replicar a aplicação da tecnologia SACA+ em diferentes níveis de ensino e contextos escolares para validar sua performance e generalização dos resultados.
- Realizar estudos que avaliem sistematicamente o impacto das diferentes ações escolares de prevenção e intervenção (como orientação, busca ativa, apoio pedagógico) sobre os estudantes identificados em risco pela tecnologia, possivelmente utilizando grupos controle, conforme sugerido no estudo do EM.
- Investigar a viabilidade e o impacto da integração da tecnologia SACA+ com outros sistemas de informação escolar para enriquecer a base de dados e potencialmente melhorar a precisão da identificação de risco.
- Aprofundar a análise de fatores demográficos e socioeconómicos, como a disparidade de evasão por gênero observada no estudo do EFII, para entender melhor os perfis de risco e adaptar as intervenções.
- Continuar otimizando os parâmetros da tecnologia para diferentes cenários de aplicação (como explorado no estudo do EM), buscando o equilíbrio ideal entre sensibilidade (identificar evadidos) e seletividade/precisão (minimizar falsos positivos), considerando o custo das intervenções escolares.

Referências

ALDEMAN, Melissa. School dropout in Central América: An overview of trends, causes consequences, and promising interventions. World Bank. 2021a. Disponível em <https://openknowledge.worldbank.org/bitstream/handle/10986/23914/School0dropout0misi ng0interventions.pdf?sequence=1&isAllowed=y> acesso 03/05/2021 12:36

ALDEMAN, Melissa. Palestra proferida no lançamento do programa Brasil da Escola. Abr. 2021b. Disponível em https://www.youtube.com/watch?v=P_kvagZ_06U acesso 03/05/2021 12:36

\

KLEIN, Ruben. Produção e Utilização de Indicadores Educacionais: Metodologia de Cálculo de Indicadores do Fluxo Escolar da Educação Básica. Revista Brasileira de Estudos Pedagógicos, v. 84, p. 107-157, 2003. Disponível em: <http://rbep.inep.gov.br/ojs3/index.php/rbep/article/view/1378/1117> . Acesso em: 3 maio 2021.

LEE, Sunbok. CHUNG, Jae Young. The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. Appl. Sci. 2019, 9, 3093; doi:10.3390/app9153093

PEREIRA JUNIOR, Edgar. Compromisso com o graduar-se, com a instituição e com o curso: estrutura fatorial e relação com a evasão. 2012. 98 f. Dissertação (Mestrado em Educação) - Programa de Pós-Graduação em Educação, Universidade Estadual de Campinas, Campinas, 2012.

THONSTAD, T. Analysing and projecting school enrolment in developing countries: a manual of methodology. Paris: Unesco, 1980. (Statistical Reports and Studies, n. 24). Disponível em: <http://unesdoc.unesco.org/images/0004/000447/044736eo.pdf> . Acesso em: 3 maio 2021

TINTO, Vincent. Dropout from higher education: a theoretical synthesis of recent research. Review of Educational Research, Washington, v. 45, n. 1, p. 89-125, Winter, 1975.