

# Universidade Federal do Espírito Santo

Resumo para a XIV Semana de Estatística do DEST

## **Reprovações escolares no ensino fundamental na cidade de Vitória-ES: modelagem preditiva**

Gabriel de Almeida Silva

Adelmo Inácio Bertolde

### **Introdução**

Reprovação no ensino médio público é um tema muito importante quando levamos em consideração o futuro dos jovens brasileiros. Segundo um estudo feito pela Unicef com dados do Inep de 2018, dos 2,6 milhões de alunos, de escolas públicas estaduais e municipais, que reprovaram no ano, mais de 1,8 milhões são do ensino fundamental, mais que o dobro dos alunos reprovados do Ensino médio.

Este é um problema grave que pode acarretar ao abandono escolar, pois, um aluno que reprova muitas vezes se sente desestimulado a continuar seus estudos, sendo que 3% dos alunos matriculados nos anos finais do ensino fundamental abandonaram a escola (Unicef, 2018).

Alguns indicadores são: 48,41% são pretos ou pardos, esse valor sendo mais do que o dobro de alunos brancos. Para alunos de classe mais baixas a repetência está fortemente relacionada a fatores socioeconômicos, o que são importantes para o sucesso escolar mentais, estruturais e de relacionamentos (Luz, 2008).

Para que esses alunos tenham uma assistência adequada e preventiva, o objetivo deste trabalho é criar um modelo preditivo estatístico, para prever se o aluno, dado suas características e desempenho no 1º trimestre do ano vigente, reprovará.

### **Dados e Metodologia**

Os dados usados para análise neste trabalho, são da secretaria de educação da cidade de Vitória - ES, de alunos de 6º a 9º ano do ensino fundamental II em escolas públicas municipais, de todos os alunos matriculados no ano de 2019.

Como o objetivo desse estudo é prever caso haja reprovação, as duas situações possíveis para o aluno são: aprovação (conclui o ano escolar com sucesso) ou reprovação (não obteve sucesso no final do ano letivo), para esta classificação final, foi utilizada como resposta a variável “situação da matrícula” de 2019 como aprovação (aprovado = 1, reprovado = 0, desistente = 0, transferido = 0), já para reprovação (aprovado = 0, reprovado = 1, desistente = 1, transferido =

0), não havia transferido em situação 2019 pois essa informação só é possível com a matrícula do ano seguinte, desistente foi incluído na análise pois o aluno ainda apresentava as notas finais.

Sobre o rendimento escolar, as variáveis mais importantes usadas são, as notas do primeiro trimestre e a nota final de cada matéria, se o aluno é público alvo da educação especial (sim=1 e não = 0), possuía registro de atendimento (não = 0 e sim = 1),

Por meio de uma análise descritiva será coletado informações e relações entre as variáveis. Pelo método de correlação de Spearman, serão retiradas possíveis variáveis muito correlatas e até mesmo concatenar informações, afim de solucionar problemas de multicolinearidade. Foram retiradas as instâncias que apresentavam dados nulos em alguma variável, e aplicados os testes de independência e de distribuição qui-quadrado ( $p\text{-valor} < 0.05$ ) para analisar a retirada de alguma classificação onde não acarretasse na distorção das distribuição das métricas.

O modelo a princípio será um modelo de predição logístico Binomial dicotômico, com a variável resposta “situação final” (aprovado, reprovado/desistente), gerado através de técnicas de Machine Learning, dividindo o banco em treino (70%) e teste (30%) e replicado diversas vezes para análises dos resultados. Para a seleção do melhor modelo será utilizado o critério de acurácia do modelo.

### **Resultados preliminares**

A base de dados contém 9520 alunos dentre eles, 687 foram reprovados em 2019 e 135 abandonaram, dentre os alunos que reprovaram, 79,76% são pardos, ressaltando que entre os alunos pardos 9,44% reprovaram ou desistiram, já os alunos brancos 6,32% dos alunos reprovaram ou desistiram e pretos 9,99% dos alunos reprovaram ou desistiram, percebe-se que dentro dos alunos negros e pardos a chance de reprovar é maior do que os alunos brancos, já alunos de raça/cor amarela, indígena e não declarado apresentam quantidades muito pequenas como 26, 25 e 24 alunos respectivamente, dentre esses apenas 4 não declarados e 1 indígena reprovou ou desistiu.

As notas dos alunos foram as variáveis com maior correlação em relação a situação final, correlacionada negativamente, ou seja, quanto menor a nota no primeiro trimestre, maior a chance de reprovar. A média de faltas dos alunos que reprovaram é 52,71 faltas, 2,29 vezes maior que a média dos alunos aprovados (27 faltas), a média dos alunos desistentes foi de 180, quando comparado com as demais mostra que muitos alunos largaram a escola muito antes, e que há flutuação nos dados em relação a faltas de alunos desistentes, levando em conta que o ano letivo tem 200 dias.

As reprovações dependem também do sexo do aluno, através do teste de independência foi concluído, dos alunos reprovados 67% eram do sexo masculino e os meninos têm 77% mais de chance de reprovar do que as meninas.

O trabalho se encontra em andamento e variáveis como escola, bairro da escola, bairro residência) notas, quantidade de pessoas dependentes ainda não foram trabalhadas profundamente, porém são variáveis importantes para o estudo.

### **Principais Referências**

FRANCESCHIN, Vanessa Lima Caldeira. A cor da reprovação: fatores associados à reprovação dos alunos do ensino médio. Educ. Pesqui., São Paulo, v. 42, n. 3, p. 773-786, jul./set. 2016. 773 DOI: <http://dx.doi.org/10.1590/S1517-9702201609149965>.

LUZ, Luciana Soares. O impacto da repetência na proficiência escolar: uma análise longitudinal do desempenho de repetentes em 2002-2003. Dissertação (Mestrado) –Universidade Federal de Minas Gerais, Belo Horizonte, 2008. Cap. 1-2.

REPROVAÇÃO, DISTORÇÃO IDADE-SÉRIE E ABANDONO ESCOLAR. Dados do Censo Escolar 2018 publicados no site da estratégia Trajetórias de Sucesso Escolar ([trajetoriaescolar.org.br](http://trajetoriaescolar.org.br)) do UNICEF e parceiros.

