

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO EDUCACIONAL
NÍVEL MESTRADO**

BRENO CÉSAR PEREIRA MONTALVÃO

**PERDA DE ALUNOS:
Sistema de Detecção Precoce na ANEAS**

Porto Alegre

2022

BRENO CÉSAR PEREIRA MONTALVAO

PERDA DE ALUNOS:

Sistema de Detecção Precoce na ANEAS

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Gestão Educacional, pelo Programa de Pós-Graduação em Gestão Educacional da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientadora: Prof^a. Dr^a. Maria Aparecida Marques da Rocha

Porto Alegre

2022

M763p	<p>Montalvão, Breno César Pereira. Perda de alunos: sistema de detecção precoce na ANEAS / por Breno César Pereira Montalvão. – Porto Alegre, 2022.</p> <p>144 f. : il. (algumas color.) ; 30 cm.</p> <p>Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Gestão Educacional, Porto Alegre, RS, 2022. Orientação: Prof. Dr. Maria Aparecida Marques da Rocha, Escola de Humanidades.</p> <p>1.Evasão escolar. 2.Estudantes do ensino fundamental. 3.Sistema de detecção precoce. 4.Escolas – Organização e administração. 5.Tecnologia educacional. I.Rocha, Maria Aparecida Marques da. II.Título.</p> <p>CDU 371.212.8 371.2</p>
-------	--

Catálogo na publicação:
Bibliotecária Carla Maria Goulart de Moraes – CRB 10/1252

Breno César Pereira Montalvão

PERDA DE ALUNOS: Sistema de Detecção Precoce na ANEAS.

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Gestão Educacional, pelo Programa de Pós-Graduação em Gestão Educacional da Universidade do Vale do Rio dos Sinos - UNISINOS

Aprovado em 2 de dezembro de 2022.

BANCA EXAMINADORA

Prof^a. Dr^a. Maria Aparecida Marques da Rocha – Universidade do Vale dos Sinos - UNISINOS

Prof^a. Dr^a. Josefina Maria Fonseca Coutinho – Centro Universitário para o Desenvolvimento do Alto Vale do Itajaí - UNIDAVI

Prof. Dr. Artur Eugênio Jacobus – Universidade do Vale dos Sinos - UNISINOS

*Aos meus amores, Magally, Maria Luiza,
Bernardo e Mateus, por me inspirarem nesta caminhada.
E, Alzira, minha mãe pelo exemplo e coragem.*

AGRADECIMENTOS

Às instituições Associação Nóbrega de Educação e Assistência Social (ANEAS) e Rede Jesuítas de Educação (RJE) pela oportunidade de crescimento e aperfeiçoamento possibilitados pela disponibilização dos dados necessários para a pesquisa.

A minha esposa, Magally Aleixo, sou grato por seu apoio e incentivo durante todo o curso. Você tem sido uma inspiração para mim.

Ao Roberto Renner e Cristiane Carmo por abrirem a porta para esta caminhada.

Ao meu irmão e amigo Brune Montalvão.

Aos colegas de trabalho Christopher Rosa Pohlmann, Marcelo Martins de Araújo e Rodrigo Rosa pelas ricas sugestões.

Aos colegas de mestrado pelo período de convivência, troca de conhecimentos e pela amizade que se formou.

A orientadora, Prof^a. Dr^a. Maria Aparecida Marques da Rocha, pelo olhar crítico e direcionamentos que possibilitaram avançar os estudos com qualidade e foco.

Aos professores, Prof. Dr. Artur Eugênio Jacobus e Prof^a. Dr^a. Josefina Maria Fonseca Coutinho, membros da banca de examinadores, obrigado por seu tempo e atenção na revisão dessa dissertação de mestrado. Agradeço todas as suas sugestões, perguntas e comentários que foram fundamentais para melhorar meu estudo.

Finalmente, àqueles que não foram mencionados diretamente, mas que fizeram parte desta etapa, meus sinceros agradecimentos.

RESUMO

A perda de alunos é um fenômeno complexo que pode ter um impacto negativo sobre o sistema educacional. Em instituições privadas, é uma importante perda de receita que pode impactar sua sustentabilidade. Está associado a fatores que incluem o não cumprimento das expectativas das pessoas quanto aos seus valores e desejos, além de causas múltiplas, relacionadas a fatores e variáveis objetivas e subjetivas. Estas estão relacionadas aos contextos econômicos, políticos, culturais e sistemas educacionais, assim como às instituições educacionais. Nesse âmbito, surge, então, a figura dos sistemas de detecção precoce de perda, do inglês *Early Detection Systems* (EDS), para a correta predição do risco de perda de alunos, tendo em vista a intervenção cada vez mais individualizada, pois partem de dados do aluno, buscando relacioná-los a fatores que indiquem, com antecedência, esse potencial de perda. Este estudo tem como objetivo analisar se o sistema de detecção precoce (EDS), a ser proposto e aplicado no contexto do ensino fundamental, anos finais dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS), oferece informações que subsidiem a detecção precoce da perda de alunos. A pesquisa busca responder à seguinte questão: o EDS aplicado no contexto dos colégios da ANEAS oferece qual resposta como instrumento de apoio aos gestores educacionais na identificação de potencial perda de alunos no ensino fundamental? Trata-se de uma abordagem mista, dos tipos documental e experimental, utilizando-se o ERP Sinergia como base de dados para a análise exploratória e para estabelecer o modelo preditivo (EDS). Os resultados deste estudo mostram que o EDS proposto é eficaz e, com a experiência e conjugação de novos atributos com maior ganho de informação, ele pode ser melhorado. Isso torna possível desenvolver uma ferramenta que ajuda a administração educacional a identificar potenciais perdas de alunos e permitir que ações sejam tomadas para retê-los.

Palavras-chave: Perda de alunos. Sistema de Detecção Precoce. *Machine Learning*. Prática da Gestão Educacional. Ensino Fundamental.

ABSTRACT

Student attrition is a complex phenomenon that can have a negative impact on the education system. In private institutions, it is an important loss of income that can impact their sustainability. It is associated with factors that include non-compliance with people's expectations regarding their values and desires, in addition to multiple causes, related to objective and subjective factors and variables. These are related to economic, political, cultural contexts and educational systems, as well as educational institutions. In this context, the figure of Early Detection Systems (EDS) emerges, for the correct prediction of the risk of losing students, with a view to increasingly individualized intervention, as they start from student data, seeking to relate them to factors that indicate, in advance, this potential for loss. This study aims to analyze whether the early detection system (EDS), to be proposed and applied in the context of elementary school, final years of the schools of the Associação Nóbrega de Educação e Assistência Social (ANEAS), offers information that subsidizes early detection the loss of students. The research seeks to answer the following question: does the EDS applied in the context of ANEAS schools offer what answer as a support tool for educational managers in identifying potential loss of students in elementary school? It is a mixed approach, of documental and experimental types, using the Sinergia ERP as a database for the exploratory analysis and to establish the predictive model (EDS). The results of this study show that the proposed EDS is effective and, with experience and the conjunction of new attributes with greater information gain, it can be improved. This makes it possible to develop a tool that helps educational administration to identify potential student losses and allow actions to be taken to retain them.

Keywords: Loss of students. Early Detection System. Machine Learning. Practice of Educational Management. Elementary School.

LISTA DE FIGURAS

Figura 1 - Modelo de cinco etapas do processo de compra do consumidor.	42
Figura 2 - Etapas entre a avaliação de alternativas e a decisão de compra	43
Figura 3 - Visão Província dos Religioso Jesuítas Brasil	55
Figura 4 – Visão geral unidades ANEAS	57
Figura 5 - Etapas do modelo de trabalho.	59
Figura 6 – <i>Framework</i> modelo preditivo	60

LISTA DE GRÁFICOS

Gráfico 1 - Número de Matrículas FUND2, ANEAS, por Período Letivo. (2018 – 2021)	67
Gráfico 2 – Número de Matrículas FUND2, ANEAS, por Unidade e Período Letivo. (2018 - 2021).....	68
Gráfico 3 – Número e Percentual de Matrículas no FUND2, ANEAS, por Ano e Período Letivo (2018 – 2021)	69
Gráfico 4 – Número de Matrículas FUND2, ANEAS, por Unidade, Ano e Período Letivo. (2018 – 2021).....	70
Gráfico 5 – Número e Percentual de Matrículas FUND2, ANEAS, por Tipo de Matrícula, Unidade e Período Letivo. (2018 - 2021).	72
Gráfico 6 – Composição População Escolar, FUND2, ANEAS, por raça/cor. (2018-2021).....	73
Gráfico 7 – Composição % População Escolar FUND 2, ANEAS, por Raça/Cor e Unidade. (2018- 2021).....	74
Gráfico 8 – % População Escolar FUND 2 ANEAS, por Gênero e Unidade. (2018-2021).....	74
Gráfico 9 – Quantidade alunos matriculados com Deficiência FUND2, ANEAS, por Tipo (2018-2021).....	75
Gráfico 10 – % População Escolar com Deficiência FUND2, ANEAS, por Unidade e Total (2018-21).....	76
Gráfico 11 – População Escolar com Registro de Deficiência FUND2, ANEAS, por Período Letivo (2018-2021)	77
Gráfico 12 – Número de Matrículas FUND2, ANEAS, por Tipo de Deficiência, Unidade e Período Letivo. (2018 – 2021)	77
Gráfico 13 – Distorção Idade-Série FUND 2, ANEAS, por ano-série. (2018 – 2021)	80
Gráfico 14 – Taxa distorção idade-série FUND2, ANEAS, por período letivo (2018-2021)	80
Gráfico 15 – Número e % de Matrículas nos Anos Finais do Ensino Fundamental Regular, ANEAS, por Distorção Idade-Série, Colégio e Ano. (2018 – 2021).....	81
Gráfico 16 – Número de Ocorrências FUND2, ANEAS, por Grupo (2018 – 2021) ...	82
Gráfico 17 - Número de Ocorrências FUND 2, ANEAS, por Unidade e Período Letivo (2018 – 2021)	83

Gráfico 18 – % de Mensalidades pagos com ou sem atraso FUND 2, ANEAS, por Período Letivo. (2018 – 2021)	84
Gráfico 19 – % de Contratos pagos com ou sem atraso FUND2, ANEAS, por Período Letivo e Colégio. (2018 – 2021).....	84
Gráfico 20 – Total de perda de alunos FUND 2, ANEAS, por colégio e período letivo. (2018-2021).....	86
Gráfico 21 – Taxa de Perda de alunos FUND2, ANEAS, por colégio e período letivo. (2018-2021).....	87
Gráfico 22 – Movimentação Acadêmica, Anchieta, por período letivo (2018-2021)..	88
Gráfico 23 – Movimentação Acadêmica FUND2, ANEAS, por período letivo (2018-2021)	89
Gráfico 24 – Taxa de Perda de alunos FUND2, ANEAS por ano/série e período letivo. (2018-2021).....	89
Gráfico 25 – Taxa de Perda de Alunos por colégio, ano/série e período letivo. (2018-2021).....	90

LISTA DE QUADROS

Quadro 1 – Trabalhos selecionados referentes aos descritores “Evasão Escolar” e “Ensino fundamental”; “Evasão Escolar” e “Ensino fundamental”	26
Quadro 2 - Etapas para implantação programa de retenção	46
Quadro 3 – Classificação de pesquisas	54
Quadro 4 – Quantidade de alunos ANEAS, por colégio. (2021)	58
Quadro 5 - Regra para classificação de perda ou retenção de alunos por “situação de matrícula”	63
Quadro 6 - Atributos da Amostra.	65
Quadro 7 – Resumo “achados”	95
Quadro 8 – Conjunto dados para treinamento e teste (2018)	98
Quadro 9 – Ranking ganho de informação atributo alvo PERDA. Base treino (2018)	98
Quadro 10 - Ranking ganho de informação atributo alvo PERDA. Base treino (2020)	100
Quadro 11 – Organização treinamento e teste do Modelo Preditivo.	101
Quadro 12 – Matriz Confusão ANEAS (2018)	102
Quadro 13 – Comparação matriz confusão EDS com hipótese apresentada	103
Quadro 14 – Análise resultados treinamento base Colégios x ANEAS	104
Quadro 11 – Organização treinamento e teste - Modelo Preditivo período Covid.	105
Quadro 15 – Matriz Confusão (2020)	105
Quadro 16 – Matriz confusão (2019 a 2021)	106

LISTA DE TABELAS

Tabela 1 - Taxa de Promoção - Dependência Administrativa Privada x Pública Ensino Fundamental.....	18
Tabela 2 - Taxa de Evasão - Dependência Administrativa Privada x Pública Ensino Fundamental	19
Tabela 3 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, Dependência Privada, região Sudeste. (2018 – 2021)	68
Tabela 4 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Ano e Dependência Administrativa Privada - Região Sudeste. (2021)	70
Tabela 5 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Sexo e Cor/Raça, Região Sudeste- 2021	73
Tabela 6 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Sexo, Região Sudeste - 2021	75
Tabela 7 – Faixa Etária Matrículas no Ensino Fundamental Anos Finais, ANEAS, por unidade (2018-2021).	78
Tabela 8 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Faixa Etária, Região Sudeste – 2021	79
Tabela 9 - Perda de alunos por Gênero, FUND 2, ANEAS – 2018-21.....	90
Tabela 10 - Perda de alunos por Raça, FUND2, ANEAS – 2018-21	91
Tabela 11 - Perda de alunos com deficiência, FUND2 ANEAS – 2018-21	91
Tabela 12 - Perda de alunos por Idade, FUND2, ANEAS – 2018-21.....	92
Tabela 13 – Idade adequada Ensino Fundamental	92
Tabela 14 - Perda de alunos por “Distorção Idade Série”, Ensino Fundamental Anos finais, ANEAS – 2018-21	93
Tabela 15 - Perda de alunos por “Tipo de Matrícula” – 2018-21	94

LISTA DE SIGLAS

ANEAS	Associação Nóbrega de Educação e Assistência Social
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CASP	Centro Administrativo São Paulo
COVID-19	<i>Corona Virus Disease</i>
EaD	Educação a Distância
EDS	<i>Early Detection Systems</i>
EJA	Educação de Jovens e Adultos
ENEM	Exame Nacional do Ensino Médio
EWS	<i>Early Warning Systems</i>
FLASC	Federação latino-Americana dos Colégios da Companhia de Jesus
FUND 2	Ensino Fundamental Regular Anos Finais (6º. ao 9º. Ano)
IDEB	Índice de Desenvolvimento da Educação Básica
IETEC	Instituto de Educação Tecnológica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LDB	Lei de Diretrizes e Bases da Educação Brasileira
MEC	Ministério da Educação e Cultura
PEC	Projeto Educativo Comum
PNAD	Pesquisa Nacional por Amostra e Domicílios
PIEA	Programa de Inclusão Educacional e Acadêmica
PPP	Projeto Político Pedagógico
RJE	Rede Jesuíta de Educação
SESI	Serviço Social da Indústria
SMS	<i>Short Message Service</i>
TCC	Trabalho de Conclusão de Curso
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

SUMÁRIO

1 INTRODUÇÃO	15
1.1 TAXAS DE FLUXO ESCOLAR NA REGIÃO SUDESTE	17
1.2 CORPO TEMÁTICO	20
1.2.1 Problema	20
1.3 OBJETIVOS.....	20
1.3.1 Objetivo geral	20
1.3.2 Objetivos específicos	21
1.3.3 Questões norteadoras	21
1.4 TRAJETÓRIA PROFISSIONAL E ACADÊMICA	21
1.5 ESTRUTURA DA PESQUISA	23
2 ESTADO DO CONHECIMENTO	25
3 REFERENCIAL TEÓRICO	30
3.1 CONTEXTUALIZANDO O SISTEMA EDUCACIONAL BRASILEIRO.....	30
3.2 EVASÃO ESCOLAR	31
3.3 COVID-19 E SUA CONTRIBUIÇÃO À EVASÃO ESCOLAR	35
3.4 PROGRAMAS DE RETENÇÃO	41
3.4.1 Processo de Compra	42
3.4.2 Fidelização	45
3.5 SISTEMAS DE DETECÇÃO PRECOCE DE EVASÃO ESCOLAR.....	47
4 ABORDAGEM METODOLÓGICA	53
4.1 MÉTODOS DE PESQUISA.....	53
4.2 PROVÍNCIA DOS RELIGIOSOS JESUÍTAS BRASIL	55
4.2.1 Conhecendo a ANEAS	56
4.2.2 Unidades educativas da ANEAS	57
4.3 CAMPO E DADOS DA PESQUISA.....	57
4.4 MÉTODO DE TRABALHO	59
5 ANÁLISE DOS DADOS	62
5.1 LEVANTAMENTO DA AMOSTRA E DEFINIÇÃO DE ATRIBUTOS.....	62
5.2 ANÁLISE EXPLORATÓRIA	66
5.2.1 Dados Demográficos	67
5.2.2 Dados Acadêmicos	81

5.2.3 Dados Financeiros	83
5.2.4 Movimentação Acadêmica.....	85
5.2.5 Principais achados da análise exploratória	95
6 EDS MODELO PREDITIVO	96
6.1 ORGANIZAÇÃO E MODELAGEM DOS DADOS	97
6.2 TREINAMENTO E TESTES.....	100
6.2.1 Treinamento e teste base (2018)	101
6.2.2 Treinamento e teste base (2020)	104
6.3 VALIDAÇÃO EDS	106
6.4 PROPOSTA DE INTERVENÇÃO.....	109
7 CONSIDERAÇÕES FINAIS.....	112
REFERÊNCIAS	118
APÊNDICE A – GANHO DE INFORMAÇÃO BASE 2018	124
APÊNDICE B – GANHO DE INFORMAÇÃO BASE 2020	131
APÊNDICE C – CLASSIFICAÇÃO EDS.....	138
APÊNDICE D – VALIDAÇÃO EDS.....	139
APÊNDICE E – ÁRVORE DE DECISÃO EDS ANEAS	141
APÊNDICE F – BI ANÁLISE PERDA DE ALUNOS	142
APÊNDICE G – BI PREDIÇÃO WEKA.....	143
ANEXO A - CARTA DE ANUÊNCIA	144

1 INTRODUÇÃO

A perda de alunos de um colégio pode ser entendida simplesmente como a transferência de seus alunos para outras instituições. A evasão escolar é compreendida como perda de alunos que iniciam, mas não concluem seus estudos saindo do sistema educacional. Em ambos os casos, nas instituições privadas constitui-se em uma importante perda de receita. É um fenômeno complexo, associado à não concretização de expectativas de pessoas e reflexo de múltiplas causas relacionadas a fatores e variáveis objetivas e subjetivas que precisam ser compreendidas no contexto socioeconômico, político e cultural, no sistema educacional e nas instituições de ensino. (FRITSCH,2015)

A desistência precoce do processo educativo é tida como uma problemática social, devido a diversas especificidades locais ou mesmo nacionais. Assim, as estratégias para manter os alunos na escola envolvem, principalmente, atrair e reter o aluno ao ambiente escolar e, ao mesmo tempo, buscar a melhoria das condições socioeducativas dos mesmos (ALMEIDA; KAIM, 2020).

O sistema educacional brasileiro compreende o conjunto de recursos para garantia à educação, esses desenvolvidos por uma série de ações e estruturas viabilizadas por instituições e entidades, públicas ou privadas. A Lei de nº 9.394, de Diretrizes e Bases da Educação Nacional, de 20 de dezembro de 1996 (LDB 9.394/96), define e regulariza o sistema de educação brasileiro em consonância à Constituição (Brasil, 1996).

Conforme o regimento escolar, documento com conjunto de regras que estrutura e estabelece todo o funcionamento e organização da instituição de ensino – nas suas esferas administrativa, didática, pedagógica e disciplinar, o sistema é dividido em Educação Básica e Educação Superior. No Brasil, a Educação Básica, constituída pelo Ensino Infantil, Ensino Fundamental e Ensino Médio, tem caráter obrigatório a qual deve ser oferecida pelo Estado como dever.

Ainda de acordo com a legislação, as instituições de ensino são classificadas em duas categorias administrativas, que incluem as públicas e privadas. Cada instituição define suas normas de gestão, levando em conta as peculiaridades existentes e deve respeitar as organizações e estrutura dos órgãos reguladores.

Então, dentro de cada categoria escolar, há a responsabilidade de constantemente desenvolver novas estratégias de aprendizagem e competências

para evitar esse problema comum na educação, a evasão escolar. Essa adversidade ocorre quando o aluno deixa de frequentar as aulas, caracterizando o abandono escolar durante o ano letivo.

Kotler e Fox (1994) destacam que, a manutenção dos alunos é determinante para as instituições educativas, uma vez que os alunos são a razão da existência destas instituições. Reter os alunos deve ser encarado como uma atividade tão importante quanto matricular novos alunos.

Posto este cenário, da educação básica como um direito constitucional e a responsabilidade das instituições definirem suas normas de gestão, um dos grandes desafios atuais dos gestores escolares passa pelo desenvolvimento de estratégias para a retenção dos alunos, incluindo conceitos de fidelização, que estabelece um relacionamento positivo entre pais, alunos e a instituição, garantindo melhores taxas de promoção e evasão, apoiando o desenvolvimento do fluxo escolar. Logo, devem entender e estruturar mecanismos que identifiquem os riscos da perda do aluno para aplicação das estratégias educacionais e de fidelização.

Alguns estudos apontam que questões acadêmicas, comportamentais, geográficas e financeiras têm relação com a decisão de pais/alunos desejarem sair ou mudar de escola. Desse modo, considera-se que mensurar e qualificar essas variáveis dentro dos colégios da Associação Nobrega de Educação e Assistência Social (ANEAS) pode oferecer um diagnóstico permanente de como está o *animus* do corpo discente, possibilitando às escolas adotarem estratégias com antecedência evitando eventuais perdas de alunos.

Para tanto, surge, então, a figura dos sistemas de detecção precoce de evasão, do inglês *Early Detection Systems* (EDS) ou *Early Warning Systems* (EWS), para a correta predição do risco de perda de alunos, tendo em vista a intervenção cada vez mais individualizada, pois partem de dados do aluno buscando relacioná-los a fatores que indiquem com antecedência este potencial de perda.

Os EDS costumam se utilizar de dados administrativos relevantes, tanto demográficos quanto de desempenho acadêmico, tais como dados pessoais (idade, gênero dentre outros), dados de educação prévia, notas e frequência escolar. Basicamente, um sistema para detecção precoce pressupõe então a existência de dados que possam ser implementados e analisados, de preferência sem o envolvimento do pessoal engajado diretamente nas instituições de ensino, facilitando consideravelmente os requisitos legais em relação às leis de proteção de dados.

Comparando os dados do ensino fundamental de dependências administrativas privadas, apresentados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), do biênio 2016-2017 com o 2017-2018 observa-se o aumento nas taxas evasão e uma diminuição da taxa de promoção, porcentagem de alunos matriculados na série seguinte em que estavam matriculados (e foram aprovados) no ano anterior, nas cidades da Região Sudeste, onde se localizam os colégios *Iócus* deste estudo.

Este estudo propõe analisar se o sistema de detecção precoce (EDS), a ser aplicado no contexto do ensino fundamental em seis colégios da ANEAS, sendo o colégio *Loyola*, em Belo Horizonte (MG), colégio *Jesuítas*, na cidade de Juiz de Fora (MG), colégio *Santo Inácio*, na cidade do Rio de Janeiro (RJ), colégio *Anchieta*, em Nova Friburgo (RJ) e, na cidade de São Paulo (SP), os colégios *São Luís* e *São Francisco Xavier (Sanfra)*, produz resultados eficientes na detecção precoce da perda de alunos, podendo ser um instrumento de apoio aos gestores educacionais.

1.1 TAXAS DE FLUXO ESCOLAR NA REGIÃO SUDESTE

Em um sistema educacional é possível avaliar a progressão de alunos a partir das taxas de transição entre séries. Para cada série existe um fluxo de entrada e saída (INEP, 2018), onde:

- a) Fluxo de entrada: alunos promovidos (alunos na série s no ano m , que estavam matriculados no ano $m - 1$ na série $s - 1$) e alunos repetentes (alunos na série s no ano m , que estavam matriculados no ano $m - 1$, na série s);
- b) Fluxo de saída: alunos promovidos à série seguinte (alunos na série $s + 1$ no ano $m + 1$, que estavam matriculados no ano m na série s), alunos repetentes (alunos na série s no ano $m + 1$, que estavam matriculados no ano m na série s) e alunos evadidos

Dessa forma, é possível o cálculo das taxas de transição entre séries, promoção e evasão. A taxa de fluxo escolar expressa o percentual de alunos promovidos (matriculados na série $s + 1$, no ano $m + 1$, em relação à matrícula total da série s no ano m), alunos repetentes (matriculados na série s no ano $m + 1$ em relação à matrícula total da série s no ano m) e alunos evadidos (alunos que estando

matriculados na série s no ano m não se encontram na matrícula da série s ou s +1 no ano m +1).

Os dados da Tabela 1 apresentam as taxas de promoção nas dependências administrativas e públicas segmentadas nos municípios e regiões geográficas em que se localizam os colégios da ANEAS. Na coluna 2017-2018, os ícones indicam se as taxas aumentaram, diminuíram ou permaneceram estáveis em comparação com o período anterior.

Tabela 1 - Taxa de Promoção - Dependência Administrativa Privada x Pública Ensino Fundamental.

Município / Região Geográfica	Dependência Administrativa	2017-2018			2016-2017		
		Anos Iniciais	Anos Finais	Total	Anos Iniciais	Anos Finais	Total
Belo Horizonte	Privada	97,3 ▼	93,7 ▼	95,8 ▼	97,4	94,0	96,0
	Pública	97,0 ▬	82,1 ▲	90,2 ▲	97,0	82,0	89,8
Juiz de Fora	Privada	97,1 ▼	93,8 ▼	95,9 ▼	97,6	94,3	96,3
	Pública	93,8 ▲	77,1 ▼	85,9 ▲	93,5	77,7	85,8
Santa Rita Sapucaí	Privada	96,4 ▼	96,0 ▼	96,2 ▼	98,2	97,2	97,7
	Pública	96,2 ▼	89,8 ▲	93,3 ▲	96,5	85,5	91,2
Minas Gerais	Privada	97,7 ▲	95,0 ▼	96,6 ▲	97,6	95,1	96,5
	Pública	96,7 ▲	83,5 ▼	90,6 ▲	96,5	83,7	90,5
Nova Friburgo	Privada	95,2 ▼	92,8 ▼	94,2 ▼	97,0	93,0	95,4
	Pública	86,8 ▲	72,3 ▼	80,1 ▼	85,3	74,4	80,2
Rio de Janeiro	Privada	94,1 ▼	91,8 ▲	93,1 ▬	94,2	91,6	93,1
	Pública	91,7 ▲	84,9 ▲	88,6 ▲	90,9	83,0	87,2
Rio de Janeiro	Privada	94,0 ▲	92,1 ▲	93,3 ▲	93,9	91,3	92,8
	Pública	87,6 ▲	78,6 ▲	83,5 ▲	86,8	77,1	82,3
São Paulo	Privada	98,0 ▲	96,5 ▲	97,4 ▲	97,9	96,0	97,1
	Pública	97,4 ▲	89,7 ▲	94,1 ▲	97,2	87,4	92,9
São Paulo	Privada	98,4 ▲	97,2 ▲	97,9 ▲	98,2	96,7	97,6
	Pública	97,1 ▲	91,9 ▲	94,8 ▲	96,8	90,6	94,0
Sudeste	Privada	96,9 ▲	95,3 ▲	96,2 ▲	96,7	94,8	95,9
	Pública	95,1 ▲	86,4 ▲	91,3 ▲	94,7	85,5	90,5
Brasil	Privada	95,9 ▬	94,5 ▲	95,4 ▲	95,9	94,1	95,2
	Pública	91,2 ▲	82,0 ▲	87,1 ▲	90,3	80,4	85,9

Fonte: Elaborada pelo autor, a partir dados INEP.

Comparando os biênios é possível verificar uma diminuição nas taxas de promoção total nas dependências privadas em todas as cidades analisadas exceto

Rio de Janeiro que manteve estável e São Paulo que apresentou aumento de 0,3 em relação ao período anterior. Por outro lado, nas dependências públicas apesar de terem taxas menores de promoção em relação às dependências privadas, registrou um aumento deste indicador no biênio 2017-18 na maioria dos municípios e regiões geográficas analisadas.

Nesse contexto, o fluxo de entrada de alunos nas instituições privadas apresenta um comportamento de diminuição entre séries, o que sugere, do ponto de vista de adesão das vagas ofertadas nas séries superiores, a necessidade de captação de novos alunos fora da instituição.

Tabela 2 - Taxa de Evasão - Dependência Administrativa Privada x Pública Ensino Fundamental

Município / Região Geográfica	Dependência Administrativa	2017-2018			2016-2017		
		Anos Iniciais	Anos Finais	Total	Anos Iniciais	Anos Finais	Total
Belo Horizonte	Privada	1,0	1,1	1,1	1,0	0,8	0,9
	Pública	0,6	2,6	1,5	0,5	2,9	1,6
Juiz de Fora	Privada	0,8	0,7	0,7	0,6	0,6	0,6
	Pública	0,9	3,3	2,0	0,8	3,6	2,1
Santa Rita Sapucaí	Privada	0,0	1,1	0,5	0,4	0,0	0,3
	Pública	0,2	2,6	1,3	0,5	3,7	2,0
Minas Gerais	Privada	0,8	0,8	0,8	0,9	0,9	0,9
	Pública	0,6	3,6	2,0	0,6	3,7	2,0
Nova Friburgo	Privada	1,0	1,0	1,0	0,3	0,9	0,5
	Pública	1,7	5,2	3,3	1,5	6,0	3,6
Rio de Janeiro	Privada	2,6	2,7	2,6	2,1	2,3	2,1
	Pública	1,6	4,0	2,8	1,6	4,7	3,1
Rio de Janeiro	Privada	2,2	2,3	2,2	1,7	2,1	1,9
	Pública	1,8	4,7	3,1	1,8	5,8	3,7
São Paulo	Privada	1,1	1,3	1,2	1,1	1,1	1,2
	Pública	0,7	2,9	1,6	0,6	3,3	1,8
São Paulo	Privada	0,7	0,9	0,8	0,7	0,8	0,8
	Pública	0,4	2,6	1,3	0,4	2,8	1,4
Sudeste	Privada	1,2	1,3	1,3	1,1	1,2	1,1
	Pública	0,7	3,4	1,9	0,7	3,7	2,1
Brasil	Privada	1,8 	1,5 	1,7 	1,5	1,5	1,5
	Pública	1,3 	4,5 	2,7 	1,3	4,6	2,8

Fonte: Elaborada pelo autor, a partir dados INEP.

Observando o fluxo de saída nos municípios levantados na Tabela 2, exceto em São Paulo, que manteve estável, a taxa de evasão nas dependências privadas aumentou em todos eles, com destaque para os municípios do Rio de Janeiro e Nova Friburgo com variações 0,5 entre os períodos avaliados.

A taxa de evasão nos períodos iniciais e finais nas instituições privadas apresentam uma variação em acréscimo, ou seja, a taxa em 2017-18 ficou superior ao período anterior. Quer dizer, nas instituições privadas ocorreu o aumento na perda de alunos, que associada a uma diminuição da taxa de promoção indica o desafio real dos gestores atuarem no fluxo escolar.

1.2 CORPO TEMÁTICO

O Corpo temático desse estudo está focado na alta administração educacional dos colégios da ANEAS, verificando a viabilidade da utilização de um Sistema de Detecção Precoce de Perda (EDS) como instrumento para gestão educacional. Trata-se de uma abordagem mista dos tipos documental e experimental.

1.2.1 Problema

A aplicação de um Sistema de Detecção Precoce de Perda (EDS) a ser desenvolvido e aplicado no contexto dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS) oferece qual resposta como instrumento de apoio aos gestores educacionais na identificação de potencial perda de alunos no ensino fundamental?

1.3 OBJETIVOS

A seguir, serão apresentados os objetivos geral, específicos e norteadores deste estudo.

1.3.1 Objetivo geral

Analisar se o sistema de detecção precoce (EDS), a ser proposto e aplicado no contexto do ensino fundamental, anos finais dos colégios da Associação Nóbrega de

Educação e Assistência Social (ANEAS), oferece informações que subsidiem a detecção precoce da perda de alunos.

1.3.2 Objetivos específicos

- a) Descrever e propor o sistema de detecção precoce de perda de alunos, por meio do conceito de EDS;
- b) Levantar os resultados dos indicadores estabelecidos no EDS no contexto da ANEAS, avaliando a predição do risco de perda de alunos;
- c) Analisar as taxas de acerto na identificação de perfis de alunos perdidos;
- d) Propor mecanismos sistêmicos e processos para implementação do EDS incluindo como rotina administrativa para apoio aos gestores na elaboração de estratégias para retenção de alunos.

1.3.3 Questões norteadoras

- a) O que são Sistemas de Detecção Precoce de Perda de Alunos do inglês *Early Detection Systems* (EDS)?
- b) Quais algoritmos são utilizados na implementação e testes do EDS?
- c) Quais os resultados do EDS aplicados no contexto dos colégios da ANEAS?
- d) Os resultados refletem o cenário real de perda de alunos?
- e) Qual o nível de acerto comparando o resultado histórico com a realidade?
- f) Quais rotinas devem ser implementadas para a correta aplicação?
- g) Como sistematizar o EDS como modelo de informação para tomada de decisão?

1.4 TRAJETÓRIA PROFISSIONAL E ACADÊMICA

Nos últimos 18 anos, construí minha carreira no mercado de Tecnologia da Informação, atuando na implantação de sistemas ERP, gestão de portfólio de projetos, gestão de serviços, vendas técnicas e relacionamento com clientes.

Em setembro de 2001, entrei na TOTVS, a partir do Projeto Formação, que capacitava profissionais do mercado na utilização do sistema ERP para posterior indicação a seus clientes ou contratação pela própria empresa. Nos anos iniciais,

trabalhei com a implantação dos módulos contábeis, financeiros, compras, estoque e faturamento.

Com os resultados adquiridos na organização, em 2008 assumi o desafio de liderar a área Portfólio de Projetos. Passei a gerenciar uma equipe de coordenadores de projetos com um portfólio anual de 80 projetos simultâneos. Além da qualidade das entregas, garantia os indicadores de resultados, como margem da operação, NPS, entre outros.

No ciclo de 15 anos na TOTVS, encerrado no final de 2016, atuei de forma direta em mais de 800 projetos de diversas complexidades. Contribuí para a consolidação da metodologia de implantação, indicando e estabelecendo artefatos que otimizavam o esforço de execução e acompanhamento.

Em 2017, passei a integrar o time do programa Sinergia ANEAS, iniciando o meu ciclo profissional no “Jesuítas Brasil”. Na empresa, fui contratado para reorganizar os vários projetos e frentes, para a unificação dos processos e sistemas das Mantenedoras ANEAS e Companhia de Jesus. Pude contribuir para a entrada em produção de todas as unidades em julho de 2017 e estabilizando no fim desse mesmo ano a operação na nova plataforma de gestão.

Atualmente, na estrutura da TI Corporativa “Jesuítas Brasil”, coordeno a TI do Centro Administrativo São Paulo, responsável pela governança e evolução dos sistemas de gestão nas unidades da “Companhia de Jesus”, “Associação Nóbrega de Educação e Ação Social” e “Fundação Fé e Alegria”.

Com o mestrado, percorri um caminho de duas vias. De um lado, assimilei a metodologia científica, de maneira a me apoiar na proposição de uma solução a um problema prático, que vivencio na rotina profissional. Em outras palavras, uma via de aprendizado que gerou maior qualificação profissional, compreendendo melhor a Gestão das Organizações Educacionais.

Do outro lado, o mestrado foi um caminho de regresso à Academia, com a qual pretendo, no futuro, estabelecer uma relação que me permita também atuar na área do ensino e formação profissional, compartilhando o repertório científico e os conhecimentos práticos acumulados na vida profissional.

Comprometido com a instituição e alinhado com minha trajetória tenho a convicção que esta pesquisa apoiará o desenvolvimento da missão e sustentabilidade de nossos colégios. Oferecendo com o resultado caminhos ou reflexões que possam

apoiar os gestores educacionais no desenvolvimento das políticas internas relacionada ao tema.

1.5 ESTRUTURA DA PESQUISA

Esta pesquisa é um estudo exploratório que pretende compreender o fenômeno da perda de alunos no ensino fundamental anos finais nos colégios da ANEAS, através de uma contextualização da estrutura educacional brasileira, com ênfase no ensino fundamental anos finais.

A pesquisa foi estruturada em seis capítulos nesta introdução, capítulo 1, aborda-se as taxas de fluxo escolar do ensino fundamental, nos municípios e regiões onde se localizam os colégios da ANEAS, nos dois últimos biênios publicados pelo INEP, o problema de pesquisa delimitado e, dele decorrentes, o objetivo geral e os objetivos específicos, o lócus da pesquisa finalizando com a apresentação profissional do autor.

O segundo capítulo apresenta o levantamento de artigos que pretende contextualizar o que tem sido feito de produção acadêmica estabelecendo diferenciais propostos nesta pesquisa.

O terceiro capítulo apresenta e se constituiu a partir da contextualização da estrutura de ensino brasileiro e problematizações da Evasão Escolar, com destaque no ensino fundamental. Descreve também o impacto da pandemia covid-19, nas instituições de ensino como uma nova variável que interfere diretamente no fluxo escolar. Aborda sobre programas de retenção no ensino superior, que possui estudos mais avançados em relação ao ensino fundamental e, complementa com o detalhamento do EDS como sistema detecção precoce de perda de alunos.

No quarto capítulo é explicitado o percurso metodológico, com o delineamento da pesquisa, o *lócus*, apresenta a metodologia empregada neste estudo, detalhando seus procedimentos relacionando os campos teóricos, que ele se insere.

Os dados empíricos deste estudo são apresentados no Capítulo 5. Organizado em seções, ele começa com o levantamento da amostra e as definições dos atributos. Segue-se uma análise exploratória dos dados agrupados por dados demográficos, acadêmicos, financeiros e de movimento acadêmico, nos períodos letivos estudados. Este capítulo é encerrado com um resumo dos resultados destas análises.

O sexto capítulo é dedicado a apresentar a proposta do Sistema de Detecção Precoce (EDS), modelo preditivo, descrevendo o percurso percorrido para sua construção, o registro dos testes e treinamentos em diferentes cenários, suas respectivas análises, validação do nível de assertividade do modelo ao aplicá-lo em outros conjuntos de dados e apresentação de uma proposta de intervenção.

Ao final, o sétimo capítulo traz considerações relacionadas ao tema da dissertação, confrontadas com os objetivos deste estudo, com o intuito de indicar se foram ou não atendidos no desenvolvimento da pesquisa, além de indicar as principais limitações e oportunidades futuras.

2 ESTADO DO CONHECIMENTO

O levantamento do estado do conhecimento foi realizado na base de dados do Portal de Periódicos da CAPES. Em 14 de abril de 2021, foram selecionados os artigos utilizando dois descritores compostos: “Evasão Escolar” AND “Ensino fundamental”; “Abandono Escolar” AND “Ensino Fundamental”.

Vale destacar que o tema perda de alunos não é muito tratado em pesquisas acadêmicas. Logo, os descritores indicados acima buscaram enriquecer a pesquisa trazendo fatores de saídas de alunos das escolas e seus impactos nas instituições e sociedade. Entretanto é importante considerar que fatores que levam à evasão não são necessariamente iguais aos fatores que levam uma família a trocar seu filho de escola.

Da primeira dupla de descritores, retornaram cento e quarenta e seis (146) resultados, sendo cento e vinte e cinco (125) artigos, treze (13) livros e oito (8) recursos textuais. E, no segundo caso, sessenta e três (63) resultados que são compostos por cinquenta e cinco (55) artigos, cinco (5) livros, dois (2) recursos textuais e uma (1) tese totalizando duzentos e seis (206) trabalhos.

Após apresentação desses resultados à orientadora, foi sugerido restringir o período dos artigos entre 2015 até a data da consulta. Desta forma, do primeiro descritor, resultaram sessenta e quatro (64) artigos e três (3) livros e, no segundo descritor, somaram trinta e três (33) artigos, três (3) livros e um (1) recurso textual totalizando noventa e oito (104) trabalhos.

A partir desse recorte, foi analisado cada trabalho com objetivo de excluir os duplicados, em cada descritor ou entre eles, que não têm autorização para divulgação e aqueles não relacionados ao tema da pesquisa.

No primeiro descritor, o artigo de Barbosa (2019) e, no segundo, os de Silva (2016) e Rocha (2018) estavam duplicados na base de dados da CAPES. Entre os descritores, retornaram os mesmos trabalhos de Silva *et al.* (2017), Filho *et al.* (2017), Filho *et al.* (2021), Santos *et al.* (2019), Leite *et al.* (2017), Assis (2015) e Negreiros (2017).

Tendo por critério os descritores definidos e os artigos cujas palavras-chave e objetivos mais se aproximam do projeto de pesquisa proposto, foram selecionados 4 (quatro) trabalhos e acrescidos 3 (quatro) da base Capes anteriores a 2015, agrupados conforme o quadro abaixo:

Quadro 1 – Trabalhos selecionados referentes aos descritores “Evasão Escolar” e “Ensino fundamental”; “Evasão Escolar” e “Ensino fundamental”.

Título	Autores	Palavras-chaves	Ano	Revista	Endereço consulta
As causas da evasão escolar: estudo de caso de uma escola pública de Ensino Fundamental no município de Acará (PA)	Marcos Jonatas Damasceno da Silva	Educação Brasileira; Educação de Jovens e Adultos; Abandono Escolar	2017	InterEspaço, 01 de março 2017, Vol.2(6), pp.367-378	https://doaj.org/article/e9d54fa7c43b439fb6c0103038eb7822
Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências.	Raimundo Barbosa Silva Filho, Ronaldo Marcos de Lima Araújo.	Ensino Fundamental – Aspectos Sociais; Evasão Escolar Brasil; Educação.	2017	Educação Por Escrito, 01 de janeiro de 2017, Vol.8(1), pp.35-48	https://doaj.org/article/3d67e97f7c8344a9be013b2c2c20ceef
Detecção precoce de estudantes em risco de evasão usando dados administrativos e aprendizagem de máquina.	José Ahirton Batista Lopes Filho, Ismar Frango Silveira.	Aprendizagem de máquina; detecção de risco de evasão; aprendizagem supervisionada; classificação; mineração de dados educacionais.	2021	Revista Ibérica de Sistemas e Tecnologias de Informação, Jan 2021, Issue E40, pp.480-495	http://search.proquest.com/docview/2483973068/?pq-origsite=primario
O abandono escolar no ensino fundamental brasileiro: alguns factores familiares.	Bayma, Hilda Freire e Assunção	Abandono escolar, Brasil	2008	Repositório Científico de Acesso Aberto de Portugal	https://estudo.geral.sib.ucp.pt/handle/10316/7512

Dificuldade de aprendizagem e o fracasso escolar.	Lisienne Navarro, Solange Gervai, Antônia Nakayama, Alice da Silva Prado.		2016	Journal of Research in Special Educational Needs, August 2016, Vol.16, pp.46-50	https://nasenjournals.onlinelibrary.wiley.com/doi/full/10.1111/1471-3802.12267
Política educacional no Brasil: educação técnica e abandono escolar	Ana Zuleima Lüscher, Rosemary Dore	Política Educacional Brasileira. Evasão, Abandono e permanência Educação Profissional. Educação Básica.	2011	Revista Brasileira de Pós-graduação, 2011, Vol.8(S1), p.S147(30)	https://rbpg.capes.gov.br/index.php/rbpg/article/download/244/411

Fonte: Elaborado pelo autor.

A seguir, serão abordados os elementos essenciais advindos dos estudos selecionados conforme quadro anterior.

O trabalho de Silva (2017) teve como objetivo identificar os principais motivos que levaram alunos de quatro turmas do ensino fundamental de uma escola pública no município de Acará, no estado do Pará, a não concluírem o ano letivo de 2015.

Nesse estudo, o autor apresenta um levantamento bibliográfico para subsidiar a elaboração do marco teórico e conceitual dos motivos do abandono escolar. Além disso, o resultado de uma pesquisa, no ano de 2016, na escola paraense, identificando quais as principais causas para estudantes não terem concluído o ano letivo de 2015.

O autor reforça, que existem vários fatores internos e externos à escola, que justificam a evasão escolar, como a necessidade de o discente trabalhar para seu próprio sustento ou de sua família, a falta de professores, a inserção do aluno na criminalidade, a falta de interesse do próprio aluno, entre outras causas. Pode-se notar, também, a partir da conclusão dessa pesquisa, que as ações para diminuição do abandono escolar não devem ser restritas à sala de aula, pelo contrário, devem ser compartilhadas através da atuação do Estado, da família, do aluno e do corpo docente.

Filho e Araújo (2017) trazem para o debate considerações sobre a evasão e o abandono escolar na educação brasileira. Compreendem que se trata de um processo muito complexo, dinâmico e cumulativo, podendo ser considerado um ato “solitário”, sem uma origem definida. Drogas, sucessivas reprovações, prostituição, falta de incentivo da família e da escola são alguns dos fatores que podem levar o educando a sair da escola.

Os autores demarcam que

fatores sociais, culturais, políticos e econômicos, bem como escolares, nos quais educadores têm colaborado a cada dia para o problema se agravar, mediante a utilização de um método didático superado ou de uma prática cristalizada como por inexperiência, acabam por desenvolver o conteúdo de forma descontextualizada e sem sentido para o aluno (SILVA FILHO; ARAÚJO, 2017, p.45).

Apontam, ainda, a importância de se detectar o interesse do aluno em sair da escola e da mediação familiar para garantir sua permanência no ensino. Isso porque, para muitos, a escola é a única fonte de informação, devido a sua condição social. Uma idade avançada na série incorreta pode ser sinônimo de múltiplas reprovações. Assim, educadores devem se esforçar para ensinar além do necessário para a conclusão do ensino e adequação de idade-série.

Filho e Silveira (2021) apresentam que as altas taxas de evasão são um problema comum em vários países, nos diversos ambientes de educação, tradicionais ou *e-learning*, e não limitadas às escolas públicas. A evasão gera um impacto negativo em todos os perfis envolvidos: estudantes, instituições e público em geral.

Os autores ressaltam, também, que sistemas de detecção precoce quanto à evasão escolar tem ganhado mais destaque, principalmente quanto à possibilidade de serem um arcabouço para a elaboração de novas políticas públicas, ao também auxiliar no melhor entendimento sobre as prováveis causas para essa evasão.

Bayma (2008) apresenta um estudo que busca averiguar fatores familiares que interferem no percurso escolar de alunos, a partir de uma amostra de 500 estudantes, de escolas públicas de Recife, dividindo-se em dois grupos, 250 que abandonaram os estudos e 250 que permaneceram na escola.

A autora conclui que fatores pessoais do aluno, como trabalho infantil e falta de interação da família com a escola, interferem diretamente no percurso dos alunos na escola. Assim, é necessário que as ações planejadas abranjam pontos intervenientes

da família, do aluno, da escola e evitem os atos que funcionam apenas como paliativos dos problemas educacionais (BAYAMA, 2008).

Navarro *et al.* (2016), verificam qual a influência das escolas no fracasso escolar do aluno. Para analisar essa questão, valeram-se de relatórios encaminhados pelas escolas para uma universidade do interior de São Paulo. Acredita-se que muitos alunos caracterizados como “alunos com dificuldade de aprendizagem” introjetaram este rótulo, bloqueando, assim, seu desenvolvimento no processo de aprendizagem, desencadeando o seu fracasso e, conseqüentemente, a evasão escolar.

Logo, os pesquisadores supracitados concluem que as escolas possuem uma força significativa na construção do fracasso e da evasão escolar devido à falta de estrutura adequada e da consciência dos profissionais da educação sobre a sua atuação, uma vez que, na atualidade, o educando necessita de um ambiente motivador e sedutor para que se desperte o interesse pelo aprender.

Lüscher e Dore (2011) apresentam análises sobre as condições que favorecem a permanência ou evasão escolar. Destacam a escassez de informações teóricas e empíricas sobre a relação da educação básica e a técnica.

De modo geral, as pesquisas recentes sobre evasão escolar apontam fatores externos e internos no que diz respeito às causas desse fenômeno.

3 REFERENCIAL TEÓRICO

3.1 CONTEXTUALIZANDO O SISTEMA EDUCACIONAL BRASILEIRO

O sistema educacional compreende o conjunto de recursos necessários para que o direito à educação seja concretizado. Esses são desenvolvidos por meio de uma série de ações e estruturas realizadas por diferentes instituições e entidades, públicas ou privadas.

A Lei nº 9.394, de Diretrizes e Bases da Educação Nacional, de 20 de dezembro de 1996 (LDB 9.394/96), define e regulariza o sistema de educação brasileiro, o qual prevê os propósitos da educação no Brasil, estabelece a sua organização, indica os órgãos administrativos responsáveis e classifica os níveis e as modalidades de ensino, entre outros aspectos incluídos na Constituição (BRASIL, 1996).

Dessa forma, de acordo com os regimentos, o sistema é dividido, em um primeiro momento, em Educação Básica e Educação Superior. No Brasil, a Educação Básica tem caráter obrigatório, a qual deve ser oferecida pelo Estado – como dever – e garantida pelos responsáveis às crianças, a fim de que concluam esse nível educacional.

Ela é constituída pelo Ensino Infantil, Ensino Fundamental e Ensino Médio. Dentro dessa classificação, ainda existem modalidades complementares para atender à demanda de alunos com necessidades especiais (Educação Especial), que desejam obter qualificação de mão de obra (Educação Profissional e Tecnológica), que não deram continuidade em seus estudos com a idade adequada (Educação de Jovens e Adultos – EJA), voltada à população indígena (Educação Indígena) e realizada à distância (EaD).

Já o Ensino Superior é facultativo para aqueles que pretendem prosseguir com os estudos universitários. Nesse caso, as modalidades incluem cursos sequenciais de bacharelado, licenciatura, tecnólogo e pós-graduações *strictu* ou *lato sensu*.

Ainda de acordo com a legislação, as instituições de ensino podem ser classificadas em duas categorias administrativas, que incluem as públicas e as privadas. As instituições públicas são mantidas e administradas pelo Poder Público, e as privadas, por pessoas físicas ou jurídicas de direito privado, podendo ser de responsabilidade dos municípios, dos estados, do Distrito Federal ou da União.

Assim, cada instituição pode definir as próprias normas de gestão, levando em conta as peculiaridades existentes em cada região e respeitando as organizações e a estrutura dos órgãos responsáveis.

3.2 EVASÃO ESCOLAR

Segundo o MEC, o conceito de evasão escolar é “a saída definitiva do curso de origem sem conclusão, ou a diferença entre ingressantes e concluintes, após uma geração completa” (MEC, 1997, p.19). Por outra forma, podemos entender como o aluno que ingressou em um curso, mas por motivos diversos não concluiu.

De acordo com Fritsch (2015, p.20):

A evasão escolar, aqui sinônimo de abandono escolar, significa desistência por qualquer motivo, exceto conclusão ou diplomação. É um fenômeno complexo, associado a não concretização de expectativas de pessoas e reflexo de múltiplas causas relacionadas a fatores e variáveis objetivas e subjetivas que precisam ser compreendidas no contexto socioeconômico, político e cultural, no sistema educacional e nas instituições de ensino.

Logo, a evasão escolar constitui-se como um ponto de atenção que se faz presente tanto em escolas públicas como privadas.

A literatura aponta que o índice maior de evasão escolar está ligado às classes sociais menos favorecidas e, dentro deste contexto, o trabalho infantil é um dos principais pontos que levam aos baixos níveis de desempenho educacionais. Esses resultados podem ser atribuídos à baixa qualidade dos serviços educacionais existentes e a ausência de uma visão crítica sobre retornos futuros.

Neri *et al.* (2008) apontam que, em 2006, conforme PNAD, cerca de 2,7% dos jovens entre 10 e 14 anos está fora da escola, contra 17,8% de jovens com 15 a 17 anos. Entre as motivações sobre a evasão apontadas por pais e pelos jovens encontram-se elementos associados à falta de demanda e deficiências de oferta.

Dentro das motivações por falta de demanda estão, primeiramente, a ausência de interesse intrínseco e, em seguida, a necessidade de trabalho e renda. Neste caso, percebe-se que a ausência de políticas públicas de ofertas de créditos educativos – como a atribuição de bolsas de estudo ou transferências de renda condicionadas –

tem uma grande importância para as repercussões negativas de alunos fora da escola. (NERI *et al.*, 2008).

Nesse sentido, Ferreira (2013) aponta que:

os motivos que levam à evasão podem ser classificados ainda de acordo com os seus fatores determinantes: (i) escola (não atrativa, autoritária, com professores despreparados, insuficiente, com ausência de motivação); (ii) aluno (desinteressado, indisciplinado, com problema de saúde, gravidez); (iii) pais ou responsáveis (não cumpridores do pátrio poder, desinteressados em relação ao destino dos filhos); (iv) social (trabalho com incompatibilidade de horário para os estudos, agressão entre os alunos, violência em relação a gangues, etc).

Já em relação aos déficits de oferta educacional, grande parte está associada à presença de deficiências físicas em estudantes, expondo os problemas de demandas específicas nas escolas para este público.

Castro *et al.* (2014) classificou os fatores escolares em dois grupos: fatores intraescolares e extraescolares:

Os fatores extraescolares são as características sociodemográficas e socioeconômicas sob as quais a escola não consegue exercer nenhum controle, mas que, em geral, influenciam no destino educacional e ocupacional dos indivíduos. Entre os fatores extraescolares, tanto a organização da sociedade, como a da família e dos próprios alunos pode ser o ponto de partida de ações direcionadas à melhoria dos resultados educacionais, pois, exercem influência direta sobre o processo de ensino-aprendizagem dos alunos e, conseqüentemente, sobre seu desempenho.

Os fatores intraescolares se referem aos aspectos e práticas escolares que influenciam no aprendizado e no sucesso escolar dos alunos atendidos pela escola. Mostram-se como características estruturais das escolas que podem ser mudadas por sua ação interna e que [...] são classificados em recursos, administração e relação da escola com a comunidade externa. Além disso, dentro da escola existem dois importantes processos que interagem para a produção do desempenho dos alunos: a gestão escolar e o ensino. (CASTRO; ALVES; NETO, 2014, p. 63).

Os riscos escolares são derivados de fatores como a cultura e o modelo de gerenciamento escolar existente, da inadequação curricular, somada à carência de trabalhos pedagógicos adequados, maiores exigências no cumprimento de metas e conteúdo, desvalorização profissional, presença de estressores no ambiente escolar, desinteresse da família no acompanhamento escolar dos alunos e outros.

Assim, essas situações promovem um ambiente com altos níveis de estresse tanto para os alunos quanto aos profissionais envolvidos na educação, causando sofrimento e, conseqüentemente, aumentando as taxas de evasão (SILVA, 2018).

Os riscos sociais do abandono escolar encontram-se nas esferas micro, meso e macrossocial. Assim, os riscos microssociais afetam diretamente o indivíduo, relacionados às suas características intrínsecas e suas relações interpessoais. Os riscos mesossociais são relativos ao ambiente de vivência, o qual o indivíduo influencia e por ele é influenciado. Por fim, os macrossociais apresentam uma grande influência no cotidiano e no papel que o indivíduo possui na sociedade. (VEIGA *et al.*, 2011).

Dessa forma, na esfera microssocial, as características individuais podem ser representadas pelo pensamento onipotente negativo, as quais geram comportamentos agressivos, impulsivos ou intolerantes às frustrações que aparecem, além de sensação de insegurança e de decepção de si mesmo.

Nas esferas meso e macrossociais, as questões de moradia, ou ainda questões de gênero, como desigualdades sociais enfrentadas pela mulher ou pela população homossexual, violência doméstica, abusos e feminicídio podem representar fatores de risco desencadeadores da evasão escolar. (VEIGA *et al.*, 2011).

A partir disso, Silva (2018) destaca que é possível identificar um caráter paradoxal no ambiente escolar, pois, ao mesmo tempo em que exigem um maior compromisso e rendimento dos seus alunos, as instituições educacionais são incapazes de garantir a estabilidade da permanência a longo prazo.

As conseqüências da evasão escolar não são apenas preocupantes para o indivíduo em si, mas também para o desenvolvimento social e econômico como um todo. Em relação ao indivíduo, o abandono promove efeitos na produtividade individual e na escola.

Nesse aspecto, a produtividade é constituída sob dois aspectos, em que o primeiro é relativo ao atraso educacional do aluno devido à inconclusão dos estudos, e outro é visto de forma mais ampla, a qual abrange o resultado da apropriação do saber, que permite a constituição do indivíduo como cidadão e sujeito histórico.

Outras conseqüências dos fatores de risco da evasão estão relacionadas à busca por prazeres imediatos ou ao aumento da autoestima, além da procura do adolescente pelo destaque na comunidade em que vive. Isso pode gerar conseqüências negativas para a sociedade, como a marginalização, criando um

contexto facilitador para o uso indevido de substâncias psicoativas. Além de promover alívio imediato para a tensão gerada pela situação estressora, pode acarretar o surgimento de transtornos de personalidades ou condutas, distúrbios depressivos e ansiedade (SILVA, 2018).

Além disso, também podem ser gerados contextos de distorção idade-série, subemprego, barateamento da mão-de-obra, dificuldade no ingresso do mercado de trabalho, estimulação da violência, prostituição, gravidez precoce e aumento dos alunos em modalidades que tentam recuperar a condição escolar, como as EJAs. Contudo, a maior consequência é o estabelecimento e a consolidação da desigualdade social, capaz de gerar uma situação de vulnerabilidade com difícil resolução. (SILVA, 2018).

A partir desse panorama, é possível perceber que, além de outras causas, o abandono escolar está associado à situação econômica e financeira individual. A partir disso, algumas estratégias governamentais, ou mesmo particulares, são importantes, a fim de evitar a evasão e manter os jovens no processo educativo. Estas oferecem apoio aos estudantes e entre os projetos podemos exemplificar “Amigos da Escola”, “Todos pela Educação” e projetos de educação do SESI. Mas, é importante destacar que a ação mais importante é dentro da escola, pois é ali que primeiro se percebe que o aluno está prestes a se evadir.

Lopes (2010) acrescenta que, para a amenização de alguns problemas referentes à evasão, é necessária uma ação firme dos poderes públicos, principalmente em relação aos gestores escolares, que precisam assegurar um bom ensino e aprendizagem.

O investimento na área da educação é de extrema importância, uma vez que ela possui efeitos diretos nas esferas sociais e econômicas do país. Isso ocorre, pois já é mostrado que um maior nível de escolaridade da população é associado a maiores salários, melhor panorama da saúde, além de diminuir a predisposição ao crime e as chances de desemprego.

Além disso, uma população mais educada resulta em aspectos econômicos positivos para o país, trazendo crescimento na produtividade de empresas e, conseqüentemente, potencializando os efeitos da globalização. Assim, é possível identificar que o processo educativo contínuo e fortalecido pode promover melhorias em diversos setores sociais (FERREIRA; OLIVEIRA, 2020).

No setor público, as perdas de estudantes também significam desperdícios sociais, acadêmicos e econômicos, pois recursos públicos são investidos em cada aluno, porém sem o devido retorno. Vale destacar que a problemática também reflete no desperdício referente a espaço físico, compra de equipamentos e aumento da ociosidade de professores e funcionários.

Assim, esse contexto promove um ciclo vicioso, já que, tanto nas instituições públicas ou privadas, uma das principais razões da evasão escolar é a falta de recursos financeiros oferecidos ao estudante para que o mesmo consiga prosseguir nos seus estudos. Do ponto de vista macroscópico, a evasão no ensino brasileiro possui uma tendência de associação com fatores socioeconômicos do país, havendo a necessidade de buscar reduzir as taxas de evasão para, com isso, evitar os desperdícios sociais e financeiros.

3.3 COVID-19 E SUA CONTRIBUIÇÃO À EVASÃO ESCOLAR

A emergência do novo coronavírus trouxe impactos negativos mundiais devido a respostas de estratégias de prevenção de contágio. Junto a estas, as estratégias de isolamento social recomendaram o fechamento de unidades escolares, afetando o funcionamento da educação.

Alguns efeitos mais críticos identificados na educação, devido à pandemia da covid-19, estão associados com o comprometimento do processo de ensino-aprendizagem e o aumento da evasão escolar.

Com isso, houve a necessidade de implementação de estratégias e formas alternativas de métodos de ensino-aprendizagem, para possibilitar a continuidade do processo educativo e evitar maiores prejuízos. Dentro destas, o ensino remoto, através do ensino a distância, se tornou um dos principais meios emergenciais para buscar suprir tal necessidade.

O emprego de atividades não presenciais constituiu-se em uma via utilizada na tentativa de minimizar as perdas causadas no processo educativo, devido às restrições e recomendações de isolamento social. Anteriormente a esse contexto, a Lei de Diretrizes e Bases da Educação Nacional (LDB) já analisava a possibilidade de realização de ensino a distância em situações emergenciais, com base no artigo 23, parágrafo 2º, que constava:

Art. 23. A educação básica poderá organizar-se em séries anuais, períodos semestrais, ciclos, alternância regular de períodos de estudos, grupos não seriados, com base na idade, na competência e em outros critérios, ou por forma diversa de organização, sempre que o interesse do processo de aprendizagem assim o recomendar.

(...)

§ 2º O calendário escolar deverá adequar-se às peculiaridades locais, inclusive climáticas e econômicas, a critério do respectivo sistema de ensino, sem com isso reduzir o número de horas letivas previsto nesta Lei (BRASIL, 2020^a).

As chances de um período longo, com atividades escolares suspensas devido às restrições da pandemia da covid-19 levaram à reorganização do calendário escolar para todas as instituições de ensino do Brasil. Essa mudança buscou garantir a continuidade das atividades escolares de forma a atender aos objetivos de aprendizagem previstos nos currículos da educação, específicos para cada nível e, ainda, cumprir as normas previstas referente a carga horária. Assim, a lei dispõe que:

A possibilidade de longa duração da suspensão das atividades escolares presenciais por conta da pandemia da covid-19 poderá acarretar: dificuldade para reposição de forma presencial da integralidade das aulas suspensas ao final do período de emergência, com o comprometimento ainda do calendário escolar de 2021 e, eventualmente, também de 2022; retrocessos do processo educacional e da aprendizagem aos estudantes submetidos a longo período sem atividades educacionais regulares, tendo em vista a indefinição do tempo de isolamento; danos estruturais e sociais para estudantes e famílias de baixa renda, como stress familiar e aumento da violência doméstica para as famílias, de modo geral; e abandono e aumento da evasão escolar (BRASIL, 2020b).

Com isso, os Conselhos de Educação de diversos estados elaboraram regulamentações para que as escolas, através da via remota, pudessem continuar suas atividades pedagógicas. Em relação à Educação Básica, o Conselho Nacional de Educação autorizou as próprias unidades de ensino da federação, estado, municípios e distritos a decidirem sobre a realização de atividades remotas em níveis do Ensino Fundamental, Médio, Técnico, Especial e de EJAs. O Ensino a Distância é tido como uma ferramenta de ensino que possibilita a autoaprendizagem, a partir da utilização de recursos didáticos organizados e apresentados nos meios de informação e comunicação. Dessa forma, as tecnologias da informação e comunicação se tornaram ferramentas importantes para o Ensino a Distância e, com isso, puderam evitar maiores prejuízos no processo de ensino e aprendizagem.

Aqui é importante destacar que existem diferenças entre ensino remoto e Educação a Distância, a qual, segundo Arruda (2020, p. 265):

A EaD envolve planejamento anterior, consideração sobre perfil de aluno e docente, desenvolvimento a médio e longo prazo de estratégias de ensino e aprendizagem que levem em consideração as dimensões síncronas e assíncronas da Ead, envolve a participação de diferentes profissionais para o desenvolvimento de produtos que tenham, além da qualidade pedagógica, qualidade estética que é elaborada por profissionais que apoiam o professor na edição de materiais diversos [...]. Já a educação remota emergencial [...] é uma mudança temporária da entrega de conteúdos curriculares para uma forma de oferta alternativa, devido à situação da crise. Envolve o uso de soluções de ensino totalmente remotas para as aulas previamente elaboradas no formato presencial, podem ser combinadas para momentos híbridos ao longo da crise, em situações do retorno parcial das aulas e quantitativo de alunos e possuem duração delimitada pelo tempo em que a crise se mantiver.

Dessa forma, a modalidade cumprida neste momento é a educação remota emergencial, que é mais conhecida simplesmente como ensino remoto. A partir disso, os educandos seguem recomendações das Secretarias de Educação, com orientações e diretrizes para o ensino remoto. Dentro dessas, são criadas estratégias específicas na tentativa de atender a todos os alunos de forma igualitária. Em relação aos alunos que necessitam de atendimento especial, a recomendação é a flexibilização de suas atividades, de forma a envolver a capacidade individual.

Como instituição social, a escola possui o importante papel de auxiliar os alunos na compreensão do atual momento vivenciado. Além disso, estratégias de ensino remoto que previnam maiores prejuízos no calendário escolar também surgem como uma alternativa para evitar o abandono escolar. Nesse sentido, é importante destacar que o uso de tecnologias a distância não tem o objetivo de substituir o ensino presencial, mas, sim, de atender a uma demanda atual e evitar que o ano letivo seja perdido.

Os problemas da continuidade do ensino-aprendizagem e a consequente evasão escolar repercutem na consolidação de dinâmicas educacionais que se manifestam entre quadros polarizados que remetem à paralisação total em oposição à continuidade das atividades por via. Entre estes, a paralisação dos processos representa o contexto mais problemático, já que a ruptura destes, devido ao contexto pandêmico, impõe certas limitações para a compreensão adequada dos conteúdos, com a nova vivência e, também, após a volta da normalidade, com os ciclos acadêmicos compactados para recuperação dos conteúdos atrasados ou mal compreendidos. Além da concentração de conteúdos no período pós pandêmico, outro fator que pode promover o aumento do abandono escolar em médio prazo é a

economia vulnerável, para a qual há a necessidade de mais esforços trabalhistas dos membros familiares (SENHORAS, 2020).

Outro fator impactante é a falta de interação social rotineira dos estudantes com seus colegas e professores, podendo acarretar o desenvolvimento de emoções que remetem à solidão ou, ainda, ao abandono. Além das dificuldades de adaptação às novas metodologias, esses fatos sociais também podem comprometer o engajamento dos alunos e, conseqüentemente, aumentar os índices de abandono escolar. Alguns dados obtidos pela Fundação Lemann mostram que o desengajamento dos alunos tem aumentado ao longo da pandemia.

De acordo com essa pesquisa, uma grande porcentagem dos pais ou responsáveis relatou desmotivação dos estudantes da educação básica para fazer as atividades da escola em casa. Em relação aos alunos do Ensino Médio, foi relatado que a desmotivação era consequência principal de dificuldades na compreensão do conteúdo e da falta de interesse. Ainda, 30% dos pais e responsáveis entrevistados na pesquisa alegaram receio de abandono da escola devido a essas objeções (FUNDAÇÃO LEMANN, 2020).

Contudo, a continuidade das atividades educacionais, através de sistemas remotos via aparelhos eletrônicos, como computador ou celular, pode preservar o comprometimento educacional, mesmo que em curto prazo. Nesse cenário, o Instituto Sonho Grande (2020) realizou uma intervenção-piloto com o objetivo de aumentar o engajamento de jovens estudantes do Ensino Médio em Tempo Integral, através do envio de *nudgebots*.

Trata-se de lembretes e sugestões enviados via mensagens de textos (SMS), que procuravam influenciar positivamente o comportamento e a tomada de decisão dos jovens, a fim de diminuir as taxas de desinteresse, de abandono e de evasão escolar.

Após a intervenção, observou-se uma redução de 77,3% nas taxas de abandono entre os alunos que recebiam as mensagens, e 86,5% destes alunos afirmam que retornarão às atividades presenciais quando as escolas reabrirem. Em relação aos alunos que não receberam as mensagens, 76% afirmam que não voltarão à escola. (FUNDAÇÃO LEMANN, 2020).

Em um contexto gerador de estresse e de sofrimento, como o atual vivenciado, o envio dos lembretes ou mensagens de acolhimento pode motivar as interações entre

os estudantes e os seus responsáveis, auxiliando na organização das rotinas de estudos.

Além disso, os *nudges* também possuem impactos positivos em matrículas, desempenho escolar, frequências de aulas assistidas, absenteísmo e na promoção de atividades que melhoram o desenvolvimento dos alunos, como aumento do tempo de atividades de leitura. Essa é uma estratégia eficiente e de baixo custo, contudo deve ser complementada com outras medidas que reforçam a retenção dos jovens no ambiente escolar (INSTITUTO SONHO GRANDE, 2020).

Em um primeiro momento, as diversas estratégias utilizadas dentro da metodologia EaD, como a transmissão adequada de conteúdos e atividades educacionais, tornaram-se um importante ponto de apoio para a solução de problemas emergenciais.

Por outro lado, a relativização dos conteúdos, em relação à qualidade do material ou à inadequada acessibilidade de alunos e professores nas plataformas educativas devido às limitações estruturais ou individuais pode representar um obstáculo ao processo educativo, necessitando de ações compensatórias após o período pandêmico (SENHORAS, 2020).

Apesar de as ferramentas *on-line* serem consideradas estratégias pertinentes para a continuidade dos estudos, principalmente no contexto de isolamento social, algumas intercorrências estruturais, ou mesmo a falta de prática e de conhecimentos específicos sobre o uso da *internet*, podem gerar limitações no ensino.

Na educação formal, a aprendizagem adequada, a partir do uso das tecnologias de informação e comunicação, depende das condições individuais e estruturais de acessibilidade. Também depende da capacitação dos professores, do nível de ensino e da idade dos alunos. Nessa linha, os resultados da aprendizagem à distância para indivíduos mais jovens, em relação à transmissão e à absorção do conteúdo, parecem não ser tão expressivos.

Para alunos do nível do ensino fundamental, a suspensão das atividades presenciais acarretou novas dificuldades, pois a mudança abrupta na rotina e nas metodologias de ensino trouxe impactos para o aprendizado dos discentes e para suas famílias, devido à sobrecarga de responsabilidade no contexto do acompanhamento do rendimento em casa (BURGESS; SIEVERTSEN, 2020).

Esse novo contexto reconfigurou a sociedade, pois todas as responsabilidades do trabalho e do estudo dos membros das famílias passaram a ser coadunadas em espaços razoavelmente reduzidos, devido à necessidade emergente.

Além disso, a crise se estende também aos docentes, que devem elaborar atividades e orientar os alunos sem a sua presença física, porém muitos não são contemplados com assistência a estrutura tecnológica necessária para as aulas remotas. Assim, utilizam seus próprios recursos, além da readequação da sua prática docente, que podem não apresentar capacidade suficiente para atender à demanda existente.

De acordo com Melo (2020), inicialmente a pandemia, os professores foram tomados por um sentimento de insegurança devido ao despreparo, tanto profissional quanto material, para o ensino remoto. Tal novidade implicaria a necessidade de adaptação radical e imediata em seus métodos de ensino. Assim, essa situação pode representar o início para alterações nas exigências curriculares da formação em docência. (PRETTO *et al.*, 2020)

Contudo, o mesmo autor identificou que a mudança no sistema de ensino resultou na maior participação dos pais ou responsáveis nas atividades escolares dos discentes, o que não ocorria expressivamente no formato de aulas presenciais. Assim, relatos mostraram um maior interesse em procurar a escola e os professores para sanar dúvidas referente às aulas ou às atividades.

Outro ponto observado foi o compartilhamento de responsabilidades referentes à educação dos jovens entre professores e responsáveis, as quais, anteriormente ao contexto atual, eram assumidas apenas pelos professores.

Apesar desse fator positivo, ainda há muito a ser feito para melhorar a condição do ensino remoto. Entretanto, essa alternativa foi importante para manter o vínculo entre os estudantes e a escola, mesmo que ainda alguns impactos futuros sejam previstos. Nesse contexto, o afastamento total dos estudantes dos espaços escolares, por diversos meses, poderia comprometer, de forma exorbitante, a qualidade da educação, muito além em comparação ao contexto atual, de implementação de iniciativas (ARRUDA, 2020).

Esses impactos sobre a educação refletem problemáticas existentes na sociedade, entre as quais, indivíduos com acesso privilegiado às tecnologias atualmente requisitadas não sofrem os mesmos efeitos da pandemia quanto a população mais vulnerável economicamente.

Dessa forma, grupos com melhores condições econômicas e maior escolarização conseguem continuar os estudos adequadamente, devido ao acesso mais facilitado a plataformas estáveis e a conteúdos de qualidade. Já para grupos em piores condições econômicas, com acesso limitado às plataformas EaD, há um maior comprometimento da continuidade dos estudos durante e até após a pandemia.

Assim, o contexto pandêmico afetou instituições com enfoque na transformação de grupos economicamente vulneráveis por meio de ações sociais e educadoras, como a ANEAS. As famílias atendidas por essas instituições geralmente têm menor poder aquisitivo, família e, em muitos casos, os responsáveis encontram-se em situação de desemprego ou subemprego, devido à pandemia.

Assim, as restrições e recomendações de isolamento afetam diretamente a continuidade das iniciativas na educação e na assistência social tanto dos jovens no processo educativo, quanto das famílias atendidas pela instituição.

De acordo com estudos realizados pela Pesquisa Nacional por Amostra e Domicílios (PNAD) – 2021, o contexto de isolamento impactará a situação econômica das famílias e, conseqüentemente, a disponibilidade de permanecer com acesso à internet. Ainda, 87,8% de indivíduos com mais de 10 anos, da região Centro-Oeste, possuem acesso à internet. Porém, destes, 92% fazem esse uso através do celular, com a principal finalidade de acesso a redes sociais ou comunicação e apenas 52% acessam a internet pelo computador (ARRUDA, 2020).

Dessa forma, o contexto da pandemia evidenciou as fragilidades do sistema socioeconômico do país e desvendou as maiores desigualdades em relação ao processo de ensino-aprendizagem.

Estas podem ser mais pronunciadas em sistemas de ensino com despreocupação em relação à inclusão social e educacional. Com isso, há a necessidade de os sistemas de ensino se comprometerem com recursos e estratégias que atendam os estudantes em sua totalidade, de forma a oferecer solução aos diversos obstáculos sociais, econômicos, digitais e culturais (UNESCO, 2020).

3.4 PROGRAMAS DE RETENÇÃO

Para compreensão do conceito de fidelização ou retenção de clientes, um dos aspectos importantes é entender o processo de compra do consumidor, ou seja, o seu comportamento diante do consumo. Para Solomon (2011, p.33), trata-se do estudo

dos processos envolvidos quando indivíduos ou grupos selecionam, compram, usam ou descartam produtos, serviços, ideias ou experiência para satisfazer necessidades e desejos.

Compreender esse comportamento permite à gestão das organizações se preparar para atuar com o consumidor nas diferentes fases do processo de compra, com informações que influenciam sua decisão de aquisição ou continuidade do consumo.

3.4.1 Processo de Compra

O processo de compra do consumidor, o qual, neste estudo, é considerado como pais e alunos, os quais estabelecem um vínculo contratual de ensino com instituições educacionais, passa por cinco etapas de atividade. Cabe destacar que, nem sempre, elas serão realizadas em sequência, podendo até pular e repetir etapas ao longo do caminho. Kotler e Keller (2011, p.166) apresentam estas etapas, conforme mostra a Figura 1.

Figura 1 - Modelo de cinco etapas do processo de compra do consumidor.



Fonte: KOTLER; KELLER (2011, p.166)

O reconhecimento do problema ou da necessidade é a etapa inicial do processo de compra, pois os consumidores assumem que têm uma necessidade não atendida (FERELL; HARTLINE, 2005). Ocorre quando o indivíduo percebe a diferença do que lhe é ideal e o seu estado atual (BLACKWELL *et al.*, 2006).

Ao perceber sua necessidade, o consumidor começa a busca por informações. Blackwell *et al.* descrevem que esta busca pode ser interna, por meio da recuperação do conhecimento na memória, ou externa, por meio da coleta de informações.

Kotler (2000) divide a busca de informações em dois níveis:

Atenção elevada: é o estado de busca mais moderado, no qual o indivíduo é mais receptivo a informações sobre o produto ou serviço.

Busca ativa de informações: é o processo mais intenso, no qual o indivíduo recorre a várias fontes de informações das quais recebe relativa influência,

de acordo com a categoria de produtos ou serviços e as características do comprador.

O tempo e o esforço que os consumidores gastam para entender um produto ou serviço dependem de três perguntas, consoante Solomon (2011, p.146) descreve:

A primeira é o nível de risco envolvido na compra. Isso significa que os consumidores, de acordo com a características do que será adquirido, pensam mais em determinados produtos e menos em outros. Os riscos não são apenas financeiros, mas podem incluir riscos sociais, risco emocional e risco pessoal. **A segunda** é a quantidade de informações ou experiências que o consumidor tem ou já teve sobre a categoria do produto, ou seja, o consumidor iniciante terá uma gama maior de instituições de ensino que, por exemplo, serão consideradas na hora matricular seus filhos. Se pais têm um problema com uma instituição ou ouvem comentários negativos de conhecidos, suas opções podem ser reduzidas. **A terceira** é o tempo disponível para a pesquisa. Consumidores pressionados pelo tempo ou com uma demanda urgente não irá dispor de tempo para realizar uma pesquisa detalhada sobre o produto ou serviços.

Na etapa de avaliação de alternativas, ocorrem as comparações entre produtos e marcas. Momento em que o consumidor julga determinado serviço e seus prestadores baseado na razão. Ele compara o que conhece dos serviços e seus respectivos fornecedores com aquilo que julga importante (BLACKWELL *et al.*, 2006).

Para Kotler (2000), o consumidor procura adquirir produtos que ofereçam os benefícios desejados, desenvolvendo uma série de crenças de marca, considerando o posicionamento dessa marca em relação a determinado atributo. Quer dizer que a avaliação dos atributos de uma marca influenciará a decisão de compra do cliente.

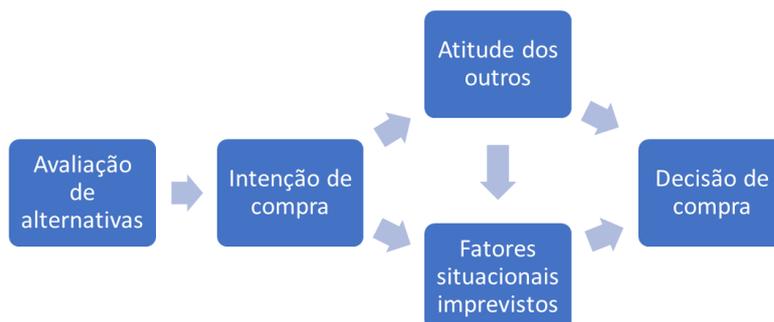
Essa etapa é o momento de escolha do consumidor da marca e ou fornecedor. Kotler e Keller (2011) descrevem que, a partir deste momento, o cliente passa a analisar cinco sub decisões: de marca, do revendedor, de quantidade, de ocasião e de forma de pagamento.

No momento da escolha, os consumidores não compram com um único tipo de regra ou diretriz. No processo de avaliação, mesmo que tenham preferência por uma marca ou fornecedor, existem dois fatores que podem interferir na intenção e decisão da compra. Kotler e Keller (2011, p.195), conforme ilustra a figura 2, definem como: a atitude dos outros e os fatores situacionais imprevistos.

O primeiro fator envolve a extensão das atitudes de terceiros na escolha do comprador para reduzir ou aumentar suas intenções, por exemplo a opinião de outros consumidores a respeito do produto escolhido. O segundo representa mudanças imprevistas, como perda de empregos, mudanças urbanas ou outras circunstâncias

imprevistas que podem aumentar ou diminuir a necessidade de compra de determinado produto.

Figura 2 - Etapas entre a avaliação de alternativas e a decisão de compra



Fonte: Kotler e Keller (2011, p.195)

A última etapa do processo de compra, comportamento pós-compra, refere-se à experiência vivida pelo consumidor após o consumo de produtos ou serviços, ou seja, a avaliação pessoal a partir de suas próprias experiências. Nessa etapa, os consumidores experimentam as sensações de satisfação ou insatisfação. (BLACKWELL *et al.*, 2006)

Conforme Schiffman e Kanuk (2009, p.393), existem três possíveis resultados da avaliação do consumidor:

- a) o desempenho real atende as expectativas levando a um sentimento neutro;
- b) o desempenho supera as expectativas, causando o que se conhece como a não confirmação negativa das expectativas o que leva a satisfação;
- c) o desempenho fica abaixo das expectativas, causando a não confirmação positiva das expectativas gerando a insatisfação.

Segundo Ferrel e Hartline (2005, p. 136), na etapa pós-compra os consumidores experimentarão quatro possíveis sentimentos:

Encantamento: quando suas expectativas foram superadas; Satisfação: quando suas expectativas foram atingidas; Insatisfação: quando suas expectativas não foram, ou foram parcialmente, atingidas; Dissonância cognitiva ou dúvida: é quando não tem certeza de que fez a escolha correta, ficando em um estado de dúvida sobre sua satisfação.

Os consumidores são o elo final na cadeia do serviço, e sua satisfação ou insatisfação afetará as vendas da empresa. Kotler e Keller (2006) explicam que um consumidor satisfeito comprará o serviço novamente e, possivelmente, até o indicará à família e amigos como sendo um serviço satisfatório.

O cliente insatisfeito, por outro lado, além de buscar a devolução do produto ou serviço, pode recorrer a ações públicas, como ir à imprensa e advogados, ou a reclamações através de redes sociais que hoje representam medo para as empresas, devido à sua disseminação de informações e ao reflexo na decisão de compra da população.

Em contrapartida, quando a satisfação alcança altos níveis, o consumidor desenvolve um vínculo afetivo com a marca, gerando um novo componente de decisão além da referência racional das informações para efetivar a compra. O resultado é um alto grau de fidelidade do consumidor (KOTLER, 2000).

3.4.2 Fidelização

A fidelidade pode ser associada com a satisfação à proporção que a lealdade e a satisfação dos consumidores aumentam, a empresa pode considerá-la como um ativo, cujo valor se amplia com o decorrer dos anos (REICHELT, 2013).

No mercado consumidor, a fidelidade é um fator chave para reter os clientes. Ferrel e Hartline (2005) destacam que a satisfação do consumidor é a chave para sua retenção. A fidelidade é uma etapa que precisa ser alcançada e não será infinita, mas quanto mais longa for a relação com o consumidor, maiores serão as chances de que ambas as partes se beneficiem. Uma instituição estruturada, com clientes relevantes por um longo período, ganha uma margem de lucro considerável para aplicar na formação de benefícios e valores para os consumidores que têm maior compromisso com a organização (MADRUGA, 2010).

Com o enfoque no segmento de ensino, na visão Kotler e Fox (1994), a organização educacional que responde ao mercado faz todos os esforços para sentir, atender e satisfazer às necessidades e aos desejos de seus consumidores e públicos.

Ainda segundo Kotler e Fox (1994), uma instituição educacional deve pesar as necessidades e preferências dos alunos e, ao mesmo tempo, preservar a reputação acadêmica e outras metas e compromissos institucionais da organização.

Cobra e Braga (2004) ressaltam que as escolas precisam estabelecer estratégias consistentes de *marketing*, para se tornarem diferenciadas no mercado, pois necessitam conquistar um posicionamento claro. A escola deve trabalhar seus valores, destacando seus atributos que a diferencia de outras escolas, com os diversos públicos com os quais ela interage, direta ou indiretamente: a)

influenciadores (familiares e amigos do aluno), b) usuários e compradores (alunos, pais e responsáveis acadêmicos). c) comprador/pagador (responsáveis financeiros).

Kotler e Fox (1994) entendem que o desafio de aumentar a retenção é de todas as áreas que compõem uma instituição escolar. Para tanto, os autores sugerem a criação de um plano de retenção, composto por oito etapas, as quais são apresentadas no Quadro 2.

Quadro 2 - Etapas para implantação programa de retenção

Etapa 1
Criar um comitê de direção da retenção, com representatividade e comprometido em trabalhar para melhorar os serviços e programas aos alunos.
Etapa 2
Estabelecer processos para análise das taxas de retenção e conhecer os fatores relacionados à transferência.
Etapa 3
Determinar as causas que levam os alunos a sair da escola, ou seja, registrar as razões apontados por alunos ou pela família para mudança de escola, bem como suas sugestões.
Etapa 4
Implantar em todas as áreas uma atitude de serviços aos alunos. Os colaboradores da instituição desde a área acadêmica até a administrativa devem estar alinhados às necessidades dos alunos.
Etapa 5
Estabelecer estratégias e planejamentos para que sejam admitidos alunos que estejam qualificados para se beneficiarem dos programas e cursos ofertados.
Etapa 6
Facilitar o acesso dos alunos e dos familiares aos ambientes do colégio. Comunicar, de forma clara, sobre os procedimentos e serviços de apoio. Além de criar no aluno um senso de pertencimento.
Etapa 7
Oferecer serviços de aconselhamento e orientação dirigidos a todos os alunos tão logo estes chegam à escola.
Etapa 8
Criar ambiente de atenção, orientado para os estudantes. Criar condições sociais e acadêmicas que encorajem o aluno a não abandonar a escola.

Fonte: Elaborado pelo autor, a partir de Kotler e Fox (1994)

No contexto deste estudo, podemos ressaltar as etapas 2 e 3 descritas acima e propostas por Kotler e Fox. A sistematização dos registros acadêmicos e instrumentos de análise de dados auxilia a gestão educacional no exame das taxas de retenção, correlacionando os principais fatores para retenção ou perda de alunos. Esses instrumentos auxiliam na detecção de potenciais perdas de alunos e no desenvolvimento de planos para fidelização.

3.5 SISTEMAS DE DETECÇÃO PRECOCE DE EVASÃO ESCOLAR

A verificação das taxas de abandono e de evasão escolar são de interesse institucional e governamental, devido a sua importância para realização de previsões socioeconômicas. Isso possibilita a elaboração de estratégias e intervenções para controle de índices referentes à qualificação profissional, ao desemprego, à valorização da mão de obra, à marginalização, entre outros. Assim, a detecção precoce de estudantes predispostos a desistir do processo educativo é fundamental para a elaboração de estratégias de retenção e de fidelização escolar (OREOPOULOS; SALVANES, 2011).

Desse modo, a identificação prévia da evasão de alunos é fundamental para evitar perdas financeiras tanto para as escolas quanto para a sociedade. A partir disso, as estratégias elaboradas e implementadas pelas organizações educacionais, para a redução das taxas de evasão escolar, devem satisfazer alguns critérios, incluindo apresentar um custo conveniente e serem direcionadas a alunos necessitados, que apresentam as maiores chances de abandonar o processo educativo. Isso implica a necessidade de identificação dos alunos a partir de dados administrativos disponíveis e a imposição de um sistema dinâmico, que corresponda às demandas e seja autoajustável (BERENS *et al.*, 2018).

Esses sistemas de detecção precoce de evasão ou *Early Detection Systems* (EDS) são capazes de gerar um percentual de risco individualizado para cada estudante, referente às chances de evasão escolar. Além disso, podem ainda ser utilizados em casos de novos ingressos ou mesmo nas transferências durante o período de um curso.

Por ser uma ferramenta considerada dinâmica, autoajustável e acessível, pode ser utilizada em sistemas públicos brasileiros, já que o método opera a partir de dados educacionais facilmente adquiridos nesses locais de ensino. Ainda, pode ser

atualizado a cada finalização de bimestre, com dados recentes dos estudantes, incluindo idade, defasagem idade-série, frequência, notas, entre outros dados demográficos e censitários (FILHO; SILVEIRA, 2021).

O sistema não é restrito apenas a uma amostra de alunos e pode ser aplicado em um contexto que avalia diferentes programas educativos, escolas e turnos, de modo a monitorar desde circunstâncias individuais até redes de ensino. Com isso, um EDS eficaz pode apresentar dados fundamentais para apoiar a elaboração de estratégias a fim de evitar o abandono escolar em instituições ou em maior escala, como novas políticas estudantis.

Outra importante influência que esses sistemas podem promover é a adequação de repasses de verbas destinadas às escolas, bem como o auxílio no monitoramento da eficiência das medidas de intervenção já exploradas.

Os EDS servem, portanto, como medida de apoio a processos operacionais, decisórios e estratégicos de instituições de ensino. Assim, permitem a avaliação de alterações em programas de ensino, de obstáculos à continuidade dos estudos, como características sociais e demográficas e monitoramento da eficiência de programas de auxílio aos estudantes em situação de vulnerabilidade (BERENS *et al.*, 2018).

Um dos primeiros modelos de sistemas de identificação precoce da evasão escolar foi proposto por Tinto (1975), no qual avaliava chances de abandono no ensino superior, através de dados que incluíam a caracterização familiar, fatores individuais e condições prévias de educação. Para o autor, esses fatores somados às características de interação dos estudantes com o ambiente de ensino no processo educativo eram as principais causas da evasão escolar. Essa interação pode ser compreendida como a integração e a adaptação acadêmica, relacionadas, ainda, ao desempenho educativo e ao desenvolvimento intelectual, além da interação entre colegas e professores.

Kember (1989) discutiu as questões relativas ao custo-benefício no processo de tomada de decisão de estudantes em continuar ou não as suas jornadas educativas, em um contexto de EaD. Assim, o autor aponta que as variáveis consideradas preditoras para a evasão poderiam não apresentar um efeito significativo direto para o abandono, mas, sim, influenciar as variáveis posteriores que afetariam diretamente o abandono escolar.

Como exemplo, tem-se a formação educacional como uma variável, mensurada através das notas do ensino médio, a qual pode não estar diretamente associada ao

abandono. Contudo, esse fator pode ter um efeito direto significativo na integração acadêmica, a qual está significativamente relacionada com índices de evasão. Portanto, ao testar um modelo, é fundamental a identificação de fatores indiretos, e isso é possibilitado pela análise de caminho, que permite também avaliar os efeitos diretos. Assim, é possível identificar os efeitos que têm maior influência sobre a variável principal analisada.

Bean e Metzner (1985) propuseram um modelo que relacionava a evasão a variáveis acadêmicas – notas e hábitos de estudo –, psicológicas – satisfação, ansiedade e outras– e ambientais – situação econômica, obrigações diárias de trabalho e outras–, além de características referentes à idade, ao gênero e ao logradouro, que poderiam estar direta ou indiretamente associadas ao processo de abandono escolar.

Nesse modelo, as variáveis de integração social podem apresentar efeitos mínimos sobre a retenção, porque, em partes, é esperado que as variáveis sociais e ambientais sejam de maior importância do que as variáveis de integração social da faculdade.

Além disso, outras variáveis ambientais, como responsabilidades familiares, podem desempenhar um papel fundamental no processo de evasão de alguns alunos. Assim, os autores sugerem que o método apresentado pode ser usado para identificar variáveis em instituições individuais e para especificar as relações entre elas, oferecendo um amplo conhecimento sobre as causas de abandono, principalmente locais.

Em contrapartida, Seidman (1996) mostra que a identificação prévia de estudantes em risco de abandonar os estudos e a constante intervenção para aumentar a retenção consistem nos principais meios para reduzir os níveis de evasão. Assim, a elaboração de novos sistemas de detecção é considerada como uma das possíveis soluções para a identificação inicial de estudantes em alto risco de se evadir do processo educativo e, com isso, há a possibilidade de adotar medidas preventivas.

É importante destacar que um EDS é basicamente um sistema projetado para alertar possíveis perigos e permitir a intervenção antes que o problema detectado se concretize. Na área da educação, o EDS atua como uma mineração de dados educacionais, em que se detectam precocemente alguns indicadores de risco de evasão escolar, porém, são necessárias condutas avaliativas adequadas sobre os resultados, a fim de gerar ações pertinentes.

Geralmente, os indicadores avaliados levam em consideração, principalmente, o desempenho escolar dos alunos, o que pode indicar o risco de abandono escolar em determinados momentos da vida acadêmica do indivíduo. Mesmo sendo uma medida muito utilizada, a identificação de fatores específicos dentro dessa variável é um trabalho difícil, já que as razões para o abandono escolar são múltiplas. Dessa forma, para a caracterização adequada dos fatores de risco, é necessário o acompanhamento desses indicadores, durante todo o desempenho escolar dos alunos, antes de abandonarem os cursos. A importância desses sistemas trouxe a necessidade de novos estudos e de investimentos para criação de novos EDS, mais assertivos e eficazes (FILHO; SILVEIRA, 2021).

Heppen e Therriault (2008) elaboraram um guia com informações sobre os fatores que contribuem para o abandono escolar de um aluno. Desse estudo, foram citados: os indicadores de alerta precoce em ambiente escolar, os sistemas de alerta precoce regionais e o papel governamental no apoio ao desenvolvimento e ao uso de EDS.

As autoras discutem que as informações contidas no guia são destinadas a apoiar educadores e a auxiliar na construção de sistemas de dados e abordagens para identificar prováveis evasões no ensino médio. Ainda, cada região, cidade ou estado deve adaptar os indicadores regionais de acordo com a sua especificidade local, econômica e social, para gerar resultados mais adequados.

Esse pode ser considerado como o primeiro passo para contemplar, de forma plena, a questão da evasão e as dificuldades encontradas pelos alunos para concluir as etapas escolares. Ainda de acordo com as autoras, os próximos passos incluiriam a identificação das características e necessidades individuais dos alunos propensos a abandonar a escola e a implementação de programas que apoiem os discentes conforme eles se encaminham para a finalização dos cursos.

Apesar da vasta literatura estrangeira sobre a temática, o Brasil também colabora com importantes conteúdos acerca da mineração de dados que mostram características e possibilidades de evasão escolar. Um desses estudos apresentou uma análise com possíveis melhorias na aplicação de mineração de dados educacionais, visando a resultados mais eficazes para a identificação de comportamentos associados à evasão escolar (RIGO *et al.*, 2012).

Dessa forma, os autores mostram a importância da realização de um amplo e, portanto, adequado mapeamento de fatores que podem estar intimamente associados

ao abandono escolar. Eles podem envolver dados históricos associados a outros mais dinâmicos, obtidos através da interação entre professores e alunos durante os semestres letivos. Além disso, também destacam a necessidade de monitoramento das ações provocadas para otimizar ações futuras, facilitar a identificação precoce e realizar intervenções pedagógicas eficientes.

Ainda em contexto nacional, outro estudo apresentou alguns fatores associados à evasão escolar e ainda descreveu o uso de um EDS, voltado a cursos de graduação a distância (RIGO *et al.*, 2014). Para a avaliação, são incluídos alguns fatores já monitorados nos sistemas utilizados, além de serem adicionados elementos referentes ao papel praticado pelo docente e a aspectos metodológicos de cada local analisado.

Como diferencial, o sistema é flexível em relação ao uso e à integração de dados adquiridos pelo processo de mediação digital, permitindo a disponibilização de diferentes ferramentas de apoio. Após a análise, os autores relataram resultados positivos, dos quais as taxas de acerto na identificação de perfis evasores encontram-se na ordem de 87%. Além disso, as taxas na redução da evasão, após a realização de intervenções, situam-se em cerca de 11%, dado que representa um importante método de detecção e prevenção do abandono escolar ou acadêmico.

De volta à realidade norte-americana, outro importante estudo forneceu um algoritmo que previa os alunos mais propensos a abandonar o ensino médio, baseando-se em informações do 9º ano do ensino fundamental (SANSONE, 2017). Segundo o autor, esse método avalia a eficácia de sistemas de identificação precoce implementados em diversos colégios, e ainda mostra que as escolas podem obter previsões mais precisas, quando exploram os dados disponíveis em associação a ferramentas de *machine learning*.

Além disso, o algoritmo seleciona cuidadosamente os critérios de adequação com base no contexto apresentado e na estrutura teórica subjacente, que incluem os objetivos das políticas e as restrições orçamentárias. Por fim, ele classifica os alunos de acordo com o risco de evasão.

Já em relação a países da América Central, como o México, o estudo de Maldonado-Ulloa *et al.* (2011) definiu algumas diretrizes para o acompanhamento da educação de alunos e desenvolveu um EDS capaz de gerar alertas, a partir de três indicadores principais. Estes incluem presença de comportamento inadequado ou problemático, absenteísmo e baixo desempenho em sala de aula. Dessa forma, a

avaliação das chances de evasão é observada através de limiares críticos que, quando ultrapassados, indicam maiores probabilidades de evasão (MALDONADO-ULLOA *et al.*, 2011).

Um estudo realizado na Dinamarca, utilizando a metodologia *machine learning*, realizou predições acerca das possibilidades de evasão escolar (SARA *et al.*, 2015). Para isso, avaliaram alunos que completaram pelo menos seis meses do ensino médio, com o objetivo de prever o abandono nos três meses subsequentes. Os dados de fontes públicas foram combinados com informações de um sistema de administração de estudos, utilizado pela maioria das escolas dinamarquesas. A amostra considerável de alunos para o estudo foi de 36299.

Após as análises, os autores obtiveram uma precisão de 93,47% e concluíram que a metodologia utilizada pode ser importante para a detecção prévia do abandono escolar, de acordo com as informações disponibilizadas pelas escolas.

Assim, para melhores resultados, é recomendada a elaboração de uma estrutura avaliativa generalista, que considere os múltiplos fatores capazes de influenciar na perda de alunos, como frequência, desempenho, características econômicas, sociais e demográficas. Dessa forma, é possível elaborar estratégias que previnam o abandono, a partir da identificação dos principais fatores que induzem a ação, diminuindo, assim, os impactos causados para a sociedade.

Logo, a partir do referencial teórico apresentado, é possível considerar que as técnicas baseadas em estatística e *machine learning*, valem-se da transformação de dados em informações ricas, possibilitando a construção de sistemas para a detecção prévia de abandono escolar.

4 ABORDAGEM METODOLÓGICA

Para Gerhardt (2009), a metodologia vai além da descrição dos procedimentos – métodos e técnicas a serem utilizados na pesquisa – indicando a escolha teórica realizada pelo pesquisador para abordar o objeto de estudo.

Este projeto trata-se de uma abordagem mista, mescla quantificação de dados e interpretação dos fenômenos, dos tipos documental e experimental. Para facilitar o entendimento do método de pesquisa escolhido, retoma-se o objetivo geral proposto: analisar se o sistema de detecção precoce (EDS), a ser proposto e aplicado no contexto do ensino fundamental, anos finais dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS), oferece informações que subsidiem a detecção precoce da perda de alunos. Ainda, tem-se como objetivos específicos:

- a) Descrever e propor o sistema de detecção precoce de perda de alunos, por meio do EDS;
- b) Levantar os resultados dos indicadores estabelecidos no EDS no contexto da ANEAS avaliando a predição do risco de perda de alunos;
- c) Analisar as taxas de acerto na identificação de perfis de alunos perdidos;
- d) Propor mecanismos sistêmicos e processos para implementação do EDS, incluindo-o como rotina administrativa para apoio aos gestores na elaboração de estratégias para retenção de alunos.

A seguir, será apresentado o resumo dos conceitos da metodologia de pesquisa definida, destacando como os autores classificam os métodos de pesquisas quanto à abordagem, à natureza, aos objetivos e aos procedimentos.

Na sequência, o campo e os dados estudados e, por fim, os métodos de pesquisas esboçados para elaboração deste trabalho, contendo suas etapas, a que cada uma se propõe, a correlação entre elas e o pilar central de sustentação.

4.1 MÉTODOS DE PESQUISA

Podemos entender que a pesquisa se constrói a partir de estudos estruturados metodologicamente para a obtenção de respostas, convincentes e reproduzíveis, ao problema proposto no projeto de pesquisa.

Segundo Gil (2002), a pesquisa é desenvolvida mediante o concurso dos conhecimentos disponíveis e a utilização cuidadosa de métodos, técnicas e outros

procedimentos científicos. Vaishnavi e Kuechler (2004) acrescentam e definem a pesquisa como sendo uma atividade que tem como objetivo contribuir para o entendimento de um fenômeno.

Gerhardt (2009) destaca que, para se fazer uma pesquisa científica, não basta o desejo do pesquisador em realizá-la; é fundamental ter o conhecimento do assunto a ser pesquisado, além de recursos humanos, materiais e financeiros.

Para Silva e Menezes (2001 *apud* SANTOS, 2004), as pesquisas podem ser classificadas pelos seguintes critérios: natureza, abordagem, objetivos e procedimentos técnicos. Estes, por sua vez, são classificados por tipos de pesquisas. A partir de tais conceitos, no Quadro 3 são apresentadas as características da pesquisa documental e logo a seguir o conceito da pesquisa experimental.

Quadro 3 – Classificação de pesquisas

Classificação das Pesquisas	Tipo de pesquisa	Descrição
Abordagem	Qualitativa	Objetiva construir o conhecimento. Nesse sentido, o processo de interpretação dos fenômenos e a atribuição de significados são básicos no processo de pesquisa.
	Quantitativa	Objetiva realizar estudos estatísticos voltados à quantificação do objeto de pesquisa.
	Mista	Mescla quantificação de dados e interpretação dos fenômenos.
Natureza	Aplicada	Objetiva gerar conhecimentos para aplicação prática dirigida à solução de problemas específicos.
Objetivos	Exploratória	Busca proporcionar maior familiaridade com o problema, para torná-lo mais explícito ou constituir hipóteses. Aprimoramento de ideias ou a descoberta de intuições.
Procedimentos	Documental	Diferente da bibliográfica, vale-se de materiais que não recebem ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa.

Fonte: Adaptado pelo autor, a partir de Silva e Menezes (2001) e Gil (2002)

A pesquisa experimental “caracteriza-se por manipular diretamente as variáveis relacionadas com o objeto de estudo. Nesse tipo de pesquisa, a manipulação das variáveis

proporciona o estudo da relação entre as causas e os efeitos de determinado fenômeno”. (CERVO; BERVIAN; SILVA, 2007, p. 63).

4.2 PROVÍNCIA DOS RELIGIOSOS JESUÍTAS BRASIL

A Província Religiosa Jesuítas Brasil (BRA), Companhia de Jesus, no aspecto canônico, organiza-se com uma liderança nacional, denominada de Provincial, que indica delegados ou secretários, como representado em vermelho na Figura 3.

Figura 3 - Visão Província dos Religioso Jesuítas Brasil



Fonte: Elaborada pelo autor.

Os superiores, delegados e secretários são responsáveis por fornecer as condições necessárias para a execução do planejamento canônico, vinculado à

missão da Companhia de Jesus. Por sua vez, eles são apoiados pelas estruturas jurídicas que tornam as ações possíveis.

Do ponto de vista legal, a BRA está estruturada em mantenedoras, representadas pelos blocos azuis da Figura 3, atuando em diversos segmentos: universidades, faculdades, ensino fundamental, assistência social, serviços, programas e projetos.

A ANEAS é uma das mantenedoras que compõem a BRA, sendo as demais: Associação Jesuítas de Educação e Assistência Social (AJEAS), Associação Antônio Vieira (ASAV), Fundação Fé e Alegria, Companhia de Jesus, Centro de Educação Técnica e Cultural (CETEC), Associação Beneficente Cultural e Educativa Padre Reus (ABEPARE) e Jesuítas Brasil.

Atualmente, a ANEAS é presidida pelo Pe. Edison de Lima SJ, com apoio da Estrutura Administrativa em São Paulo, liderada pelo Ir. Eudson Ramos SJ Administrador Provincial BRA.

4.2.1 Conhecendo a ANEAS

A Associação Nóbrega de Educação e Assistência Social (ANEAS) tem como missão promover a pessoa humana de maneira integral, à luz da fé e buscando a justiça, em diálogo com diversas expressões culturais e religiosas.

Como propósito, a associação garante os meios para o fortalecimento da Rede Jesuíta de Educação e Assistência Social e busca ser reconhecida como entidade de vanguarda na formação integral da pessoa humana.

Além disso, almeja ser vista como referência em gestão, e ser percebida como uma instituição onde o reconhecimento e o desenvolvimento das potencialidades individuais contribuem para a excelência no ambiente de trabalho (PEC, 2018).

Do ponto de vista de Estado, trata-se de uma instituição de direito privado e sem fins lucrativos, filantrópica, de natureza educativa, cultural, assistencial e beneficente. Possui o Certificado de Entidade Beneficente de Assistência Social nas áreas de educação e de assistência social e desenvolve o Programa de Inclusão Educacional e Acadêmica (PIEA), por meio de bolsas de estudo para educação básica.

Na estrutura jurídica da ANEAS, encontram-se filiais educativas do ensino infantil, fundamental, médio, técnico, Educação para Jovens e Adultos (EJA), centros sociais, museus, gráfica e mosteiro.

4.2.2 Unidades educativas da ANEAS

A figura 4 apresenta as principais unidades que compõem a ANEAS, seguida por uma descrição dos seis colégios *locus* desta pesquisa.

Figura 4 – Visão geral unidades ANEAS



Fonte: Elaborada pelo autor.

Os colégios da ANEAS com ensino fundamental anos finais (FUND2) estudados nesta pesquisa estão localizados na Região Sudeste do país, sendo: o “Colégio Loyola” em Belo Horizonte (MG), “Colégio Jesuítas” na cidade de Juiz de Fora (MG), “Colégio Santo Inácio” na cidade do Rio de Janeiro (RJ), “Colégio Anchieta” em Nova Friburgo (RJ) e na cidade de São Paulo (SP) os colégios “São Luís” e “São Francisco Xavier”.

4.3 CAMPO E DADOS DA PESQUISA

Quanto ao campo da pesquisa, as instituições escolhidas apresentam aspectos similares. Por exemplo, o público atendido é majoritariamente das classes A e B, baseadas no Projeto Educativo Comum (PEC) da Rede Jesuíta de Educação (RJE),

e, exceto o colégio *Anchieta*, estão localizados nas grandes capitais da Região Sudeste.

Segundo a pesquisa do INEP, no período avaliado (2016-2017 vs. 2017-2018), o Sudeste é uma das regiões do Brasil que mostrou um aumento nas taxas de evasão escolar e uma diminuição nas taxas de promoção escolar nas Dependências Administrativas Privadas.

Como instrumento de coleta de dados, as informações dos alunos foram extraídas do banco de dados do ERP ANEAS, em operação desde o segundo semestre de 2017, com registros unificados de todas os colégios. Ele contém todas as informações administrativas das unidades educacionais estudadas, incluindo detalhes de matrícula dos estudantes, responsáveis financeiros e acadêmicos, registros de notas, faltas e ocorrências, movimento acadêmico e situação financeira da mensalidade. Em 2021, as unidades totalizavam 12.177 alunos. Isto inclui os alunos do ensino fundamental e médio. (Quadro 4).

Quadro 4 – Quantidade de alunos ANEAS, por colégio. (2021)

Unidade	Cidade	Estado	Número de alunos
Anchieta	Nova Friburgo	RJ	889
Jesuítas	Juiz de Fora	MG	2.033
Loyola	Belo Horizonte	MG	2.396
Santo Inácio	Rio de Janeiro	RJ	3.533
São Francisco Xavier	São Paulo	SP	1.123
São Luís	São Paulo	SP	2.203
Total			12.177

Fonte: Elaborado pelo autor, a partir de métricas enviadas para Censo (2020).

Após a definição dos atributos a serem estudados, os dados empíricos foram extraídos, organizados e analisados, considerando os períodos escolares de 2018 a 2021. Em 2020 e 2021, o Brasil foi atingido pela pandemia da covid-19. Embora esse cenário seja atípico em relação aos períodos escolares anteriores, neste estudo foi considerado importante destacar as variáveis que precisavam ser observadas no novo contexto.

A partir dos resultados da análise dos dados empíricos e da maior compreensão dos atributos relacionados aos dados demográficos, acadêmicos e financeiros de cada instituição foi estabelecido o conjunto de dados para treinar e

testar o modelo preditivo (EDS), baseado em árvore de decisão. A seção seguinte deste capítulo detalha o método de trabalho adotado para construção do Modelo Preditivo.

4.4 MÉTODO DE TRABALHO

O método de trabalho para o projeto de pesquisa foi constituído por quatro etapas, compostas por passos que forjaram os resultados do projeto. No conjunto das ações, há uma correlação entre as ações que permite retomar etapas anteriores, amadurecendo e qualificando o desenvolvimento da pesquisa.

Como pilar central, o referencial teórico foi o gatilho e o balizador para a evolução das etapas, conforme apresentado na Figura 5.

Figura 5 - Etapas do modelo de trabalho.



Fonte: Elaborada pelo autor.

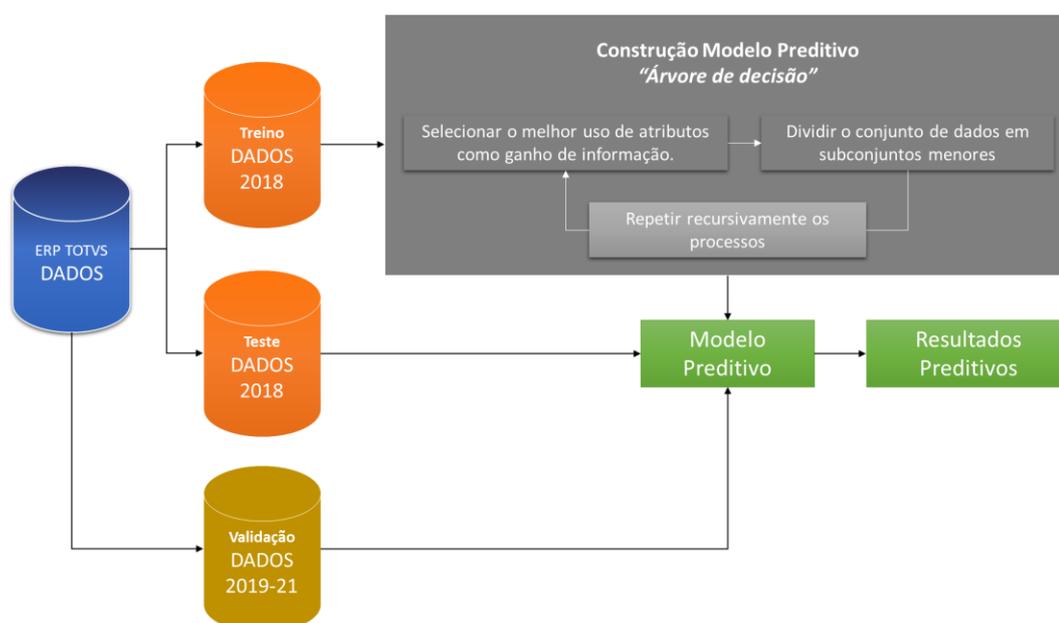
A seguir, apresentam-se as etapas constitutivas do processo de pesquisa, seguidas de uma breve explicação sobre o significado e a importância do processo junto ao contexto metodológico.

Construção do Modelo Preditivo: Para a construção do modelo preditivo, foram definidas as variáveis conforme referências bibliográficas, destacando que essas variáveis dizem mais respeito ao abandono e menos à decisão de troca de escola, e o conjunto de dados extraídos do ERP da ANEAS. Elas foram organizadas para análise exploratória. Após a compreensão dos dados, a partir do contexto da pesquisa, foram definidos: o atributo foco e os atributos da categoria y, para montagem

do modelo preditivo. As amostras foram categorizadas em duas classes: perdas e retidos.

Conforme figura 6, os dados foram organizados em conjuntos menores para treino, teste e validação do modelo preditivo. Os conjuntos de dados foram treinados e testados no *Weka*, um *software* de mineração de dados gratuito e de código aberto, desenvolvido pela Universidade de Waikato, localizada na Nova Zelândia. A ferramenta reúne uma coleção de algoritmos de aprendizagem de máquinas para tarefas de mineração de dados, como pré-processamento, classificação, regressão, agrupamento, associação e visualização de dados.

Figura 6 – *Framework* modelo preditivo



Fonte: Adaptada de FILHO; SILVEIRA (2021).

Os resultados dos testes foram analisados na própria plataforma *Weka*, que gera a matriz de confusão como resultado do cálculo do algoritmo definido. Vale ressaltar que este projeto não se propõe a estudar ou comparar algoritmos de classificação de dados. Na fase de indução da árvore de decisão, *Decision Forest*, foi definido o algoritmo de classificação J48¹ reconhecido por estudos pelo poder de predição.

¹ O algoritmo J48 é um algoritmo de árvore de decisão que pode ser usado para criar modelos de árvore de decisão. Ele utiliza tecnologia *greedy* para induzir árvores de decisão para posterior classificação, e é projetado para trabalhar com dados de treinamento, bem como dados ainda não classificados.

O conjunto de dados definido para a construção do modelo preditivo foi o de 2018. Ou seja, a árvore de decisão do modelo foi induzida pelo conhecimento gerado por esses dados, e os demais conjuntos de dados (2019, 2020 e 2021) foram utilizados na validação do nível de acurácia do modelo estabelecido. Um ponto a considerar é que 2020 e 2021 foram anos completamente atípicos, em função da pandemia, quando fatores incomuns atuaram, para além dos tradicionais.

Tendo em vista a amplitude de informações, o treino, o teste e a definição do modelo foram focados na ANEAS e eventualmente, realizadas simulações com dados exclusivos dos colégios, para comparação dos resultados de um modelo preditivo genérico, que atenda a todas as unidades, com modelos preditivos específicos.

Aplicação do Modelo Preditivo: definido o modelo preditivo, ele foi aplicado a um novo conjunto de informações não treinadas, com dados de 2019, 2020 e 2021. Foram organizados os resultados a cada inclusão de conjunto de dados inseridos no modelo preditivo definido. Da mesma forma que na fase de construção, inicialmente as validações foram feitas com dados consolidados na ANEAS, incluindo eventualmente resultados específicos por colégio, para enriquecer a análise.

Análise de Resultados: com os resultados organizados, foram realizadas as análises de acurácia do modelo preditivo, verificando se os valores apontados representam uma realidade. Ou seja, foram verificadas as taxas de acerto na identificação de perfis evasores entre os indicadores. Para tal, utilizou-se a Matriz de Confusão². Em seguida, foram identificados os atributos com maior ganho de informação disponíveis no ERP, para correta predição de perda de alunos. Assim, foram indicadas eventuais melhorias no pré-requisito e no *framework* do EDS.

Conclusão: esta etapa destina-se a apresentar as conclusões referentes ao desenvolvimento da pesquisa e à proposição de um eventual modelo sistêmico que suporte a consolidação do modelo preditivo definido, constituindo-se uma ferramenta de detecção precoce de perda de alunos para apoio aos gestores nas tomadas de decisão, focadas na retenção de alunos.

² A Matriz de Confusão é uma tabela em que facilmente identificamos todos os quatro tipos de classificação do modelo de classificação binária. Com ela, podemos calcular valores, como acurácia, especificidade, sensibilidade etc. Ela é uma característica que trata da qualidade da classificação realizada, mostrando quais foram os casos em que o algoritmo se confundiu na classificação. A diagonal principal da matriz apresenta as instâncias classificadas corretamente para cada classe e corresponde à precisão do classificador

5 ANÁLISE DOS DADOS

Esta etapa da pesquisa tem como objetivo definir, organizar e analisar os dados empíricos do estudo para predição da perda de alunos. Através destas análises, respeitando os objetivos específicos do projeto, serão estabelecidos os critérios iniciais para a construção do modelo EDS que será aplicado em uma etapa posterior.

A seguir, são apresentados os seguintes tópicos: levantamento da amostra e definição de atributos, análise exploratória dos dados, considerando os demográficos, acadêmicos, financeiros e as movimentações acadêmicas. Finalizando o capítulo, será apresentado o resumo dos achados da pesquisa que contribuíram para a fase de estruturação do EDS.

5.1 LEVANTAMENTO DA AMOSTRA E DEFINIÇÃO DE ATRIBUTOS

Segundo Freitas (2016), são várias as metodologias para se compreender a evasão, as quais estão ligadas aos conceitos adotados, aos dados disponíveis e à equação a ser utilizada para se obter um índice de evasão. Ainda, segundo o autor, existem diferentes percursos e maneiras de se mensurar a ocorrência da evasão, entretanto dependerá dos dados e das situações particulares das instituições.

Para entender o contexto da perda de alunos e a construção do sistema EDS, foram utilizados os dados do ERP Sinergia¹, um sistema de planejamento de recursos empresariais implementado na ANEAS em julho de 2017. No ERP, os dados são estruturados e atualizados seguindo conceitos e processos similares em todas as escolas onde a pesquisa foi realizada, permitindo a consolidação da análise.

Os dados foram extraídos do ERP, através de relatórios administrativos e consultas SQL² em seu banco de dados, contemplando o registro de todos os alunos

¹ ERP Sinergia e a sigla de Enterprise Resource Planning. É um sistema de gestão, implantado de forma integrada nas unidades da ANEAS, que permite acesso fácil, integrado e confiável aos dados da empresa. A partir das informações levantadas pelo software, é possível fazer diagnósticos aprofundados sobre as medidas necessárias para reduzir custos e aumentar a produtividade.

² SQL significa “*Structured Query Language*”, ou “Linguagem de Consulta Estruturada”, em português. Resumidamente, é uma linguagem de programação para lidar com banco de dados relacional (baseado em tabelas). Foi criado para que vários desenvolvedores pudessem acessar e modificar dados de uma empresa simultaneamente, de maneira descomplicada e unificada.

do ensino fundamental dos seis colégios da ANEAS, do 6º ao 9º ano, no período entre 2018 e 2021.

Para a análise inicial e visualização dos dados, optou-se pelos softwares Excel® e PowerBI®³ da Microsoft. Isso porque, assim, foi possível unir a facilidade do Excel para sanear os dados brutos à destreza do Power BI para montagem de visualizações interativas e recursos de *business intelligence*⁴.

A fim de compreender uma variável específica, é importante estabelecer uma conexão com sua variável contraditória. Isso significa que, se quisermos saber por que os alunos deixam a escola, é igualmente importante saber por que eles permanecem na escola.

Para a categorização dos dados coletados, utilizou-se o status "situação de matrícula". A partir deste status, foi estabelecida a regra para identificar os alunos que saíram (perda) ou permaneceram na escola (retidos). O Quadro 5 apresenta as "situações de matrícula" relacionadas à perda ou retenção dos alunos.

Quadro 5 - Regra para classificação de perda ou retenção de alunos por “situação de matrícula”

		TOTAL DE ALUNOS	
		Perda	Retidos
Situação da Matrícula	Cancelado		Aprovado
	Desistente		Concluído
	Transferência Externa		Reprovado
	Transferência entre Unidade		

Fonte: Elaborado pelo autor, a partir das Regras de Negócio Sinergia ANEAS (2017)

Sobre as “situações de matrícula” relacionadas à perda de alunos, entende-se:

Cancelado: interrupção da validade de matrícula mediante solicitação dos pais e/ou responsáveis antes do início do período letivo;

³ O Power BI é uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente envolventes e interativas.

⁴ *Business Intelligence* pode ser descrito como um processo orientado pela tecnologia para analisar dados e apresentar informações acionáveis para ajudar executivos, gerentes e outros usuários finais corporativos a tomar decisões de negócios bem-informadas.

Desistente: caso haja desistência do curso antes do início das aulas, entende-se que não houve efetiva prestação de serviços da instituição de ensino;

Transferência externa: saída do aluno mediante solicitação dos pais e/ou responsáveis. Esse status também é utilizado, para identificar o aluno que sai da instituição de ensino por motivo de “Intercâmbio”;

Transferência entre unidades: A transferência de alunos entre unidades indica que os pais e/ou responsáveis solicitaram transferência, para outra instituição de ensino da Companhia de Jesus.

Para critério de análise, a “transferência entre unidades” foi considerada como perda de aluno, pois, mesmo que ele permaneça em outra unidade da Companhia de Jesus, no contexto do colégio é uma perda e, conseqüente, diminuição de receita.

Sobre as situações da matrícula relacionadas à retenção:

Aprovado: A condição é aplicável aos alunos que alcançaram, os critérios mínimos para conclusão satisfatória da etapa de ensino que estava cursando, estando apto para ser matriculado na etapa seguinte no próximo ano letivo;

Concluído: É considerado aluno concluinte aquele que completou, com emissão de certificado, a etapa que estava cursando (Ensino Fundamental, Ensino Médio ou Ensino Técnico) na instituição de ensino;

Kotler e Fox (2014) descrevem a importância de as instituições de ensino conhecerem os registros históricos e analisarem os fatores relacionados às transferências de alunos. Nesse contexto, os atributos definidos para prever o risco de perda de alunos, além de correlacionados a saídas dos estudantes, também devem ser obtidos por meio dos dados administrativos, na ANEAS, registrados no ERP.

Conforme os estudos levantados, fatores de grande importância na análise da perda ou retenção de estudantes são o foco em suas características demográficas, bem como seu desempenho acadêmico, sua frequência e a relação contratual com a instituição.

Segundo Tinto (1993), os indivíduos ingressam no ambiente escolar com uma variedade de atributos pessoais (sexo, raça/cor e habilidades físicas), recursos financeiros, disposições (preferências intelectuais, sociais e políticas) e diferentes tipos de experiências e realizações educacionais. Cada atributo é postulado como tendo um impacto direto na evasão.

O Quadro 6 apresenta a nomenclatura definida para os atributos selecionados, possíveis de consolidação entre as unidades, com suas respectivas descrições. As informações são baseadas nos dados sistematizados, mantidos pelas unidades no sistema ERP Sinergia, que serão utilizados para a análise exploratória e para o desenvolvimento do Modelo Preditivo (EDS).

Quadro 6 - Atributos da Amostra

Atributo	Detalhamento
Y_PERDA	Rótulo se o aluno saiu ou não do colégio (0 - retido - e 1 - Perda).
SEXO_ALUNO	Referente a sexo biológico (F - feminino - ou M - masculino).
RAÇA_ALUNO	Raça/cor aluno (0 - não consta, 1 - branca, 2 - preta, 3 - parda, 4 - amarela, 5 - indígena ou 6 - não declarada).
POSSUI_DEFICIÊNCIA	Rótulo se o aluno possui deficiência ou não (0 - não e 1 - sim)
IDADE_NOLETIVO	Idade do aluno no período letivo
DISTORÇÃO_IDADE	Rótulo se o aluno possui distorção idade-série no período letivo (0 - não e 1 - sim)
ATRASO	Rótulo se o aluno possui pelo menos uma parcela paga com mais de 30 dias do vencimento. (0 - não e 1 - sim)
QDE_PARCELAS	Número de parcelas pagas em atraso conforme critério ATRASO
OCORRÊNCIA	Rótulo se o aluno recebeu pelo menos uma ocorrência no período letivo dos tipos “notificação escolar”, “notificação disciplinar” e “orientação escolar”. (0 - não e 1 - sim)
QDE_OCORRÊNCIAS	Quantidade de ocorrências conforme atributo OCORRÊNCIA.
FALTA_MAT	Taxa de frequência na matéria Matemática do aluno, no referido período letivo
FALTA_PORT	Taxa de frequência na matéria Português do aluno, no referido período letivo
TIPO_MATRICULA	Rótulo tipo de matrícula (0 – rematrícula e 1 – matrícula)
TIPO_ENSINO	Nome dado ao atributo tipo de ensino (Ensino Fundamental).
SÉRIE	Número da série em que o aluno se encontra (de 6 a 9).
TURNO	Nome dado ao atributo turno (matutino, vespertino, noturno e integral).
PLETIVO	Período letivo
TX_PERDA	Porcentagem de perda de alunos na escola.

Fonte: Elaborado pelo autor

5.2 ANÁLISE EXPLORATÓRIA

Após a coleta e o saneamento dos dados, foram estruturadas tabelas fatos⁵ e tabelas dimensões⁶ para modelagem dos dados e construções de painéis no Power BI para as análises iniciais das informações geradas, de modo a detectar seus subconjuntos e formular as primeiras hipóteses, a partir desses painéis.

Segundo Bantenero (1991), o ponto central da análise exploratória de dados é que o uso de múltiplas representações de um conjunto de dados propicia o desenvolvimento de novos conhecimentos e ideias. A análise exploratória de dados nos fornece um extenso repertório de métodos para um estudo detalhado dos dados, antes de adaptá-los.

Cabe lembrar que as análises foram organizadas por dados demográficos (total de alunos matriculados, raça/cor, sexo, idade e deficiência), dados acadêmicos (notas, ausências e ocorrências), dados financeiros (parcelas pagas em atraso) e os dados da movimentação acadêmica entre os períodos letivos, verificando as incidências por variáveis, no contexto de perda ou retenção de alunos.

Para complementar as análises iniciais, este projeto também utilizou o banco de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep, 2021), que torna públicos os dados do Censo Escolar⁷ em seu site desde 1995. Os dados extraídos foram limitados ao período de 2018 a 2021, anos finais do ensino fundamental regular (FUND2), da Região Sudeste e, às vezes, quando disponíveis, os dados por dependência privada.

A fim de facilitar a compreensão dos resultados apresentados nas próximas seções, quando a sigla ANEAS for utilizada, ela se referirá à análise que consolida os dados das seis escolas que foram *lócus* da pesquisa.

⁵ Tabela de fatos é a tabela dominante de um esquema de modelagem tipo estrela (Star Schema), criado pelo Dr. Ralph Kimball, em um modelo multidimensional, e tem como característica principal a presença de dados altamente redundantes para se obter um melhor desempenho.

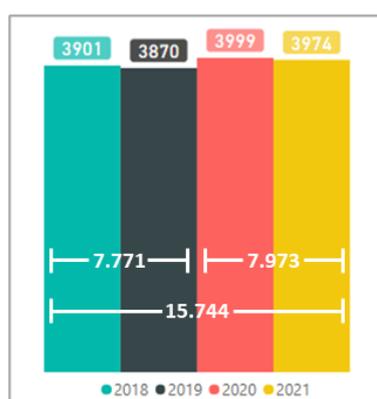
⁶ A tabela dimensão é aquela que vai auxiliar a tabela fato com dados complementares ou explicativos; ela possui informações que não se repetem.

⁷ O Censo Escolar é o principal instrumento de coleta de informações da educação básica e a mais importante pesquisa estatística educacional brasileira. É coordenado pelo Inep e realizado em regime de colaboração entre as secretarias estaduais e municipais de educação e com a participação de todas as escolas públicas e privadas do país.

5.2.1 Dados Demográficos

Nesta seção, será apresentada a análise dos dados demográficos obtidos no universo pesquisado. O Gráfico 1 mostra a quantidade total de matrículas (15.744) realizadas no 6º ao 9º ano do ensino fundamental, na ANEAS, organizadas nos períodos letivos estudados (2018-21). Em 2020, foi atingido o maior número de alunos matriculados, com 3.999, enquanto em 2019 o menor, com 3.870.

Gráfico 1 - Número de Matrículas FUND2, ANEAS, por Período Letivo. (2018 – 2021)



Fonte: Elaborado pelo autor.

É importante notar que, em 2020 e 2021, o número de estudantes na ANEAS foi maior em comparação com o biênio anterior. Nesses dois anos, o Brasil experimentou uma pandemia causada pela covid-19, que gerou um grande impacto social e econômico no país. Famílias e escolas foram diretamente afetadas em sua renda, no trabalho e na dinâmica de prestação de serviços, como descrito no Capítulo 7: Covid-19 e sua contribuição para o abandono escolar.

De certa forma, em relação à população educacional do FUND2, pode-se deduzir que, em geral, os colégios absorveram os impactos resultantes dos efeitos sociais causados pela covid-19, pelo fato de que o número de alunos matriculados na fase pandêmica (7.973) foi 2,5% maior do que no biênio pré-pandêmico (2018- 2019), com 7.771 alunos.

Quando comparado ao desempenho das escolas privadas na Região Sudeste, as escolas jesuítas mostraram um melhor crescimento no número de matrículas. Nas escolas da ANEAS, o crescimento foi de 2,5%, enquanto, neste mesmo período, o

crescimento das matrículas nas escolas da região foi de 0,8% (20/21 – 1.715.440 em relação 18/19 – 1.701.297) (conforme Tabela 3).

Tabela 3 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, Dependência Privada, Região Sudeste (2018 – 2021)

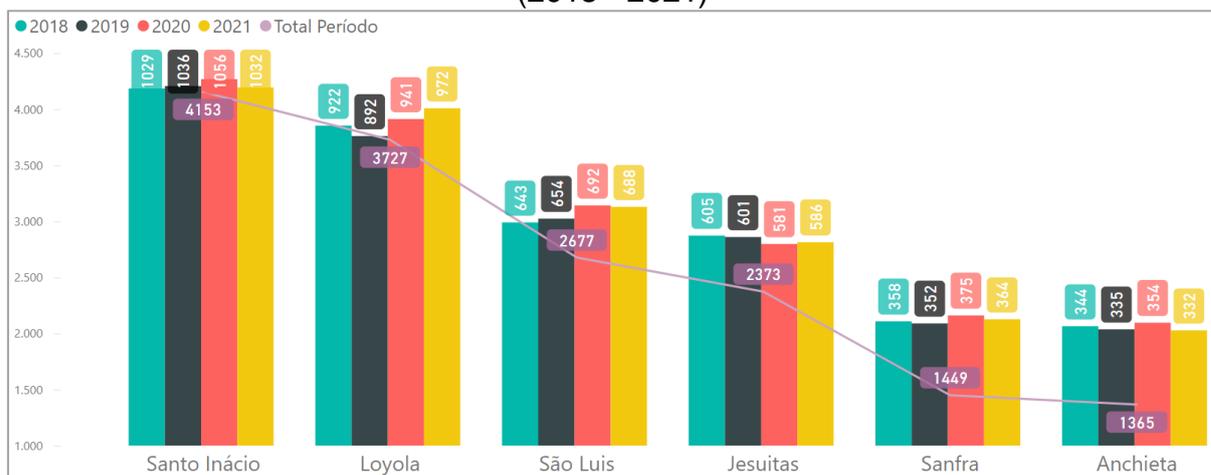
Ano	Ano								Biênios		
	Total	6º ano		7º ano		8º ano		9º ano			
2018	845.641	227.450	26,9%	207.503	24,5%	207.034	24,5%	203.654	24,1%	1.701.297	49,8%
2019	855.656	232.545	27,5%	220.902	26,1%	201.498	23,8%	200.711	23,7%		
2020	862.956	235.264	27,8%	222.853	26,4%	211.108	25,0%	193.731	22,9%	1.715.440	50,2%
2021	852.484	223.162	26,4%	219.642	26,0%	209.031	24,7%	200.649	23,7%		
Total	3.416.737	918.421	26,9%	870.900	25,5%	828.671	24,3%	798.745	23,4%		

Fonte: Censo Escolar (Inep 2018 -21)

O Gráfico 2 mostra a quantidade total de matrículas realizadas do 6º ao 9º ano do ensino fundamental, distribuídas nas unidades educacionais da ANEAS e seus respectivos períodos letivos. Os colégios possuem faixas distintas quanto ao volume de alunos.

No total do período estudado, o Santo Inácio (RJ) e o Loyola (BH) assumiram a faixa mais alta, com 4.153 e 3.727 alunos matriculados, respectivamente; na faixa intermediária, com próximos a 2,5 mil alunos, São Luís (SP) e Jesuítas (MG) e, como menor volume de alunos, Sanfra (SP) e Anchieta (RJ), na faixa de 1,4 mil alunos. Cabe, portanto, sinalizar que os colégios de maior porte possuem até três vezes mais alunos do que os de menor porte.

Gráfico 2 – Número de Matrículas FUND2, ANEAS, por Unidade e Período Letivo. (2018 - 2021)



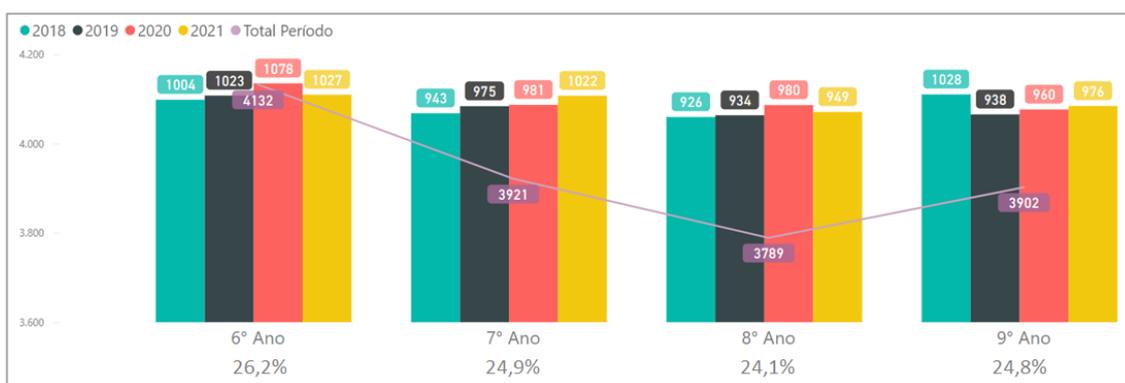
Fonte: Elaborado pelo autor.

Apesar de haver características similares entre os colégios jesuítas, o volume de alunos pode sugerir maior complexidade de gestão escolar e, conseqüentemente, um maior esforço para garantia dos processos administrativos, como a atualização e a manutenção das rotinas sistêmicas.

Segundo o Inep (2021), o número de alunos é um dos fatores para determinar a complexidade da gestão escolar. Para estabelecer os níveis de complexidade, o Inep avalia não apenas o tamanho da escola – que está associado ao volume de matrículas –, mas também o número de turnos (manhã, tarde ou período integral) e o número de etapas de educação fornecidas (ensino fundamental, ensino médio e EJA).

Com relação aos estudantes matriculados, o gráfico 3 ilustra a distribuição das vagas preenchidas nos períodos escolares de 2018 a 2021, agrupadas por anos finais do ensino fundamental. Ao analisar a linha "Total Período", que soma as vagas preenchidas nas séries históricas por ano/série, nota-se uma queda nos alunos matriculados entre 6º ano ao 8º ano, ocorrendo um aumento no 9º ano (3.902), mas ainda inferior ao 6º e 7º ano. Na soma do período avaliado, as unidades têm uma porcentagem mais concentrada de alunos matriculados no primeiro ano da série, 6º ano (26,2%), e um pequeno equilíbrio (faixa de 24%) nos outros anos.

Gráfico 3 – Número e Percentual de Matrículas no FUND2, ANEAS, por Ano e Período Letivo (2018 – 2021)



Fonte: Elaborado pelo autor.

No período letivo mais recente (2021), em comparação com a análise anterior que considerou todos os períodos letivos, apesar de haver um maior equilíbrio entre os anos/séries, também concentram-se as matrículas no 6º ano (1.027), com 25,8% do total de matrículas no período. Para os demais anos, os percentuais foram: 7º ano

(1.022), 25,7%; 8º ano, 949 (23,9%) e, com menor percentual, o 9º ano, com 976 (24,6%).

Conforme dados do INEP, na Tabela 4, os colégios da Região Sudeste, no ano de 2021, apresentam concentração de matrículas entre ano-série próximas aos colégios da ANEAS, entretanto com percentuais diferentes e o 9º ano com menor percentual (23,5%) que as escolas aqui estudadas.

Tabela 4 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Ano e Dependência Administrativa Privada - Região Sudeste (2021)

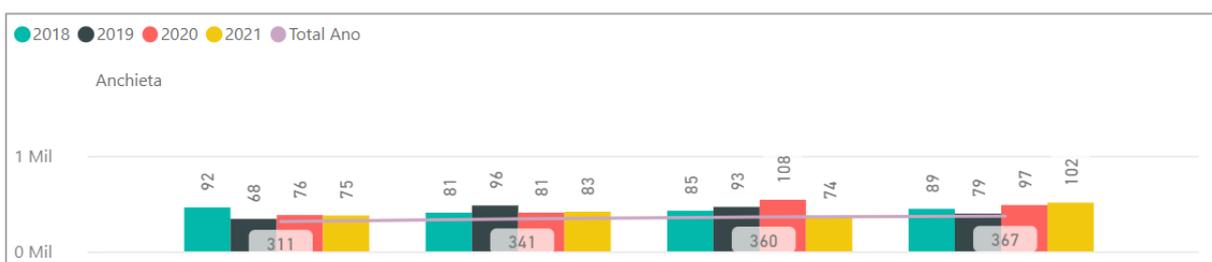
Total	Ano			
	6º ano	7º ano	8º ano	9º ano
852.484	223.162	219.642	209.031	200.649
	26,2%	25,8%	24,5%	23,5%

Fonte: Censo Escolar (Inep 2021)

O Gráfico 4 apresenta uma visão semelhante ao Gráfico 3, incluindo a dimensão colégio. Explorando os dados por colégio, a curva da linha "Total Ano", entre o ano inicial e final, do Loyola, Sanfra e Santo Inácio segue o mesmo padrão de redução de matrículas entre o 6º ano e 9º.ano, da curva "Total Período" apresentada no gráfico anterior.

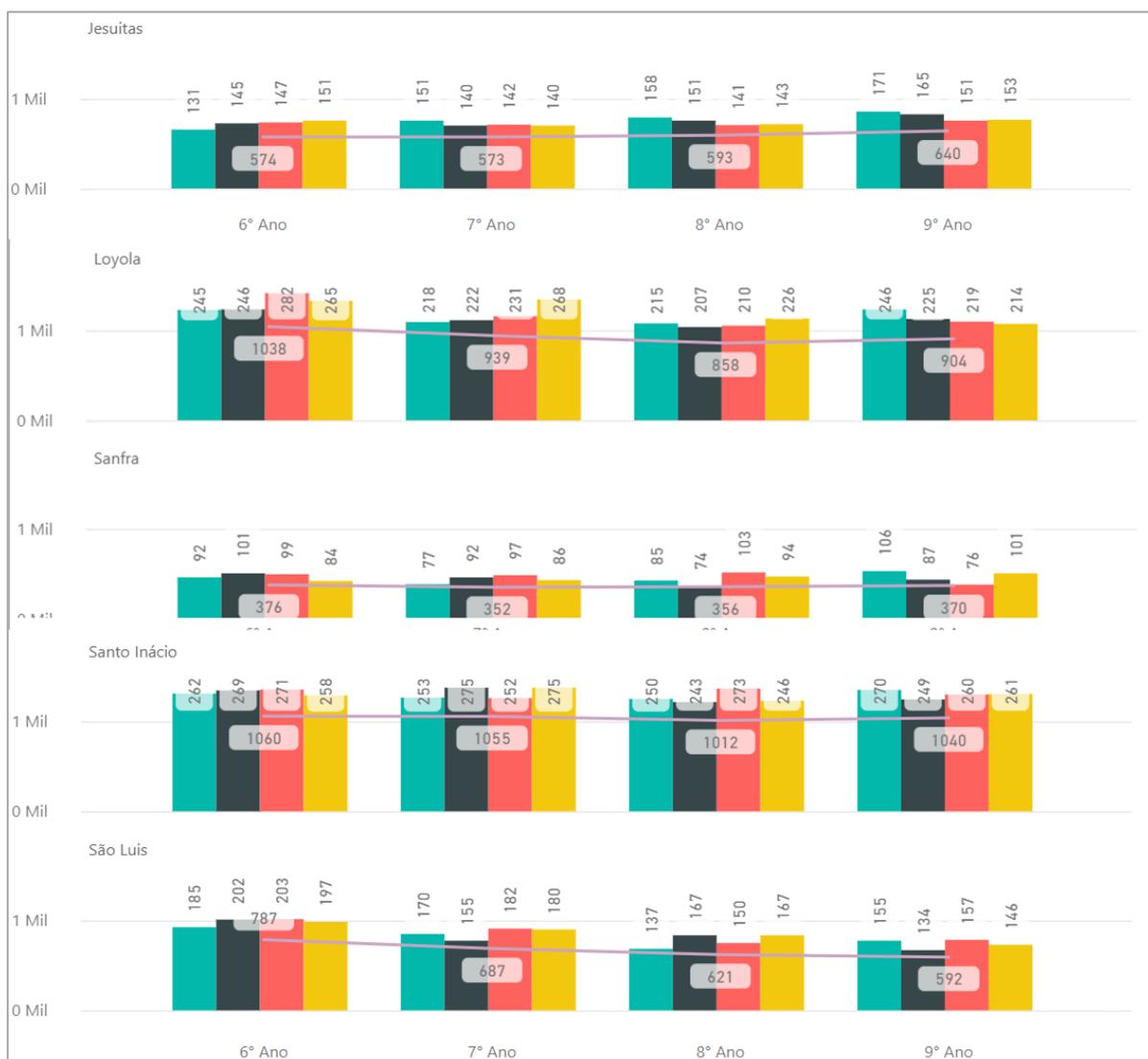
Entretanto, os colégios Anchieta e Jesuítas exibem um percentual maior de matrículas nos últimos ano/série, e o *São Luís*, com uma queda constante e mais acentuada entre o 6º e o 9º ano. Isso sugere que, embora tenha havido um menor volume global de matrículas entre os anos/séries, a partir do 6º ano, essa redução não foi uniforme em todos os colégios; ao contrário, alguns experimentaram quedas mais drásticas do que outros.

Gráfico 4 – Número de Matrículas FUND2, ANEAS, por Unidade, Ano e Período Letivo. (2018 – 2021)



Fonte: Elaborado pelo autor.

Gráfico 4 – (Continuação) Número de Matrículas FUND2 ANEAS, por Unidade, Ano e Período Letivo. (2018 – 2021)



Fonte: Elaborado pelo autor.

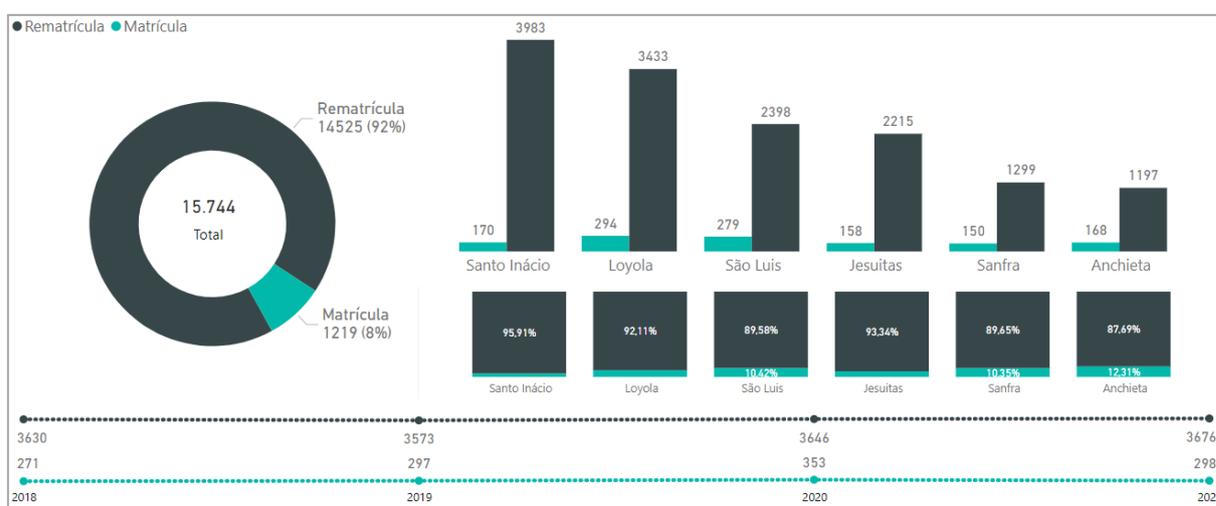
O Gráfico 5 mostra o volume de matrículas de alunos novos ou veteranos. Os dados revelam que, na ANEAS, 92% dos alunos inscritos no período 2018-21 são veteranos, ou seja, alunos promovidos do ano anterior ou repetentes que mantiveram seu vínculo com o colégio. Portanto, pode ser evidenciado um percentual de fidelização, tendo em vista que a maioria dos estudantes mantêm seu vínculo com os colégios em sucessivos períodos letivos.

Quando olhamos os dados por unidade, *Santo Inácio*, *Jesuítas* e *Loyola* têm a maior porcentagem de rematrículas, com 95,91%, 93,24% e 92,11%, respectivamente. Enquanto *Anchieta*, *São Luís* e *Sanfra* têm as menores porcentagens de rematrículas, com 87,69%, 89,58% e 89,65%, respectivamente.

Por outro lado, o percentual de novos alunos matriculados no mesmo período foi de 8% (1.219 matriculados). Na série histórica, 2020 exibiu o maior fluxo de novos alunos, em relação ao total de inscrições (matrícula mais rematrícula), com 8,8% (353 matrículas do total de 3.999 inscrições). Esse fato pode justificar o crescimento da população educacional no período pandêmico, descrito anteriormente.

Em 2018, o número de novos alunos na ANEAS atingiu seu ponto mais baixo nos quatro anos estudados, com uma matrícula total de 271 novos alunos. Esse número representa 6,9% do das inscrições no ano.

Gráfico 5 – Número e Percentual de Matrículas FUND2, ANEAS, por Tipo de Matrícula, Unidade e Período Letivo. (2018 - 2021).



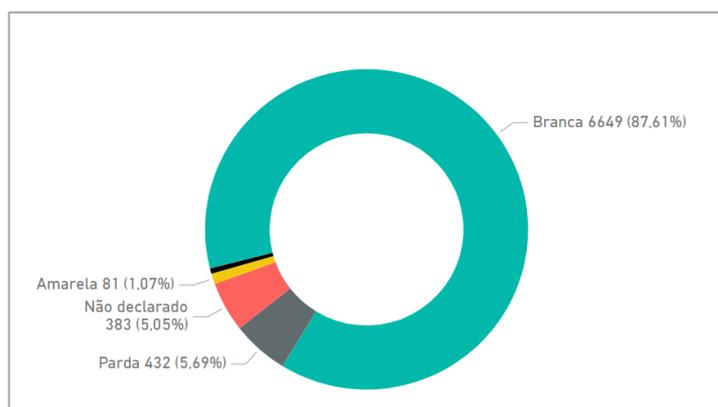
Fonte: Elaborado pelo autor.

A composição da população escolar por raça/cor no período analisado é mostrada no Gráfico 6. Na soma de todas as escolas, "branca" representa 87,61%, "parda" 5,69%, "amarela" 1,07% e menos de 1% "preta" e "indígena", com 42 e 2 alunos respectivamente. Ao mesmo tempo, 5,05% dos alunos não tiveram sua cor/raça declarada.

Ao contrastar esses dados com a realidade brasileira, pode-se inferir que alguns que não declaram sua raça/cor poderiam se identificar como pretos ou pardos, mesmo assim, a alta representação dos brancos e a sub-representação dos pardos/pretos chamam a atenção.

De acordo com dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2019, 42,7% dos brasileiros se declararam como brancos, 46,8% como pardos, 9,4% como pretos e 1,1% como amarelos ou indígenas.

Gráfico 6 – Composição População Escolar, FUND2, ANEAS, por raça/cor. (2018-2021)



Fonte: Elaborado pelo autor

Segundo o Censo Escolar 2021 (tabela 5), o percentual de alunos matriculados como brancos no ensino fundamental anos finais na Região Sudeste – onde estão localizados os colégios pesquisados – é de 45,3%, e os percentuais declarados como pretos correspondem a 3,6%, e de pardos, 31,5%. Ressalta-se, desse modo, a baixa representação de pardos/negros na ANEAS.

Tabela 5 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Sexo e Cor/Raça, Região Sudeste- 2021

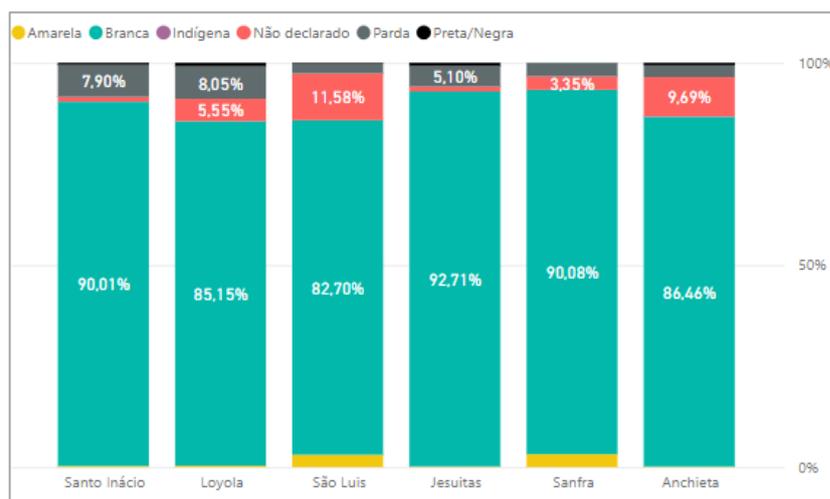
Gênero	Total	Raça/cor											
		Não Declarada		Branca		Preta		Parda		Amarela		Indígena	
Feminino	2.237.356	431.054	19,3%	1.022.107	45,7%	79.022	3,5%	695.013	31,1%	7.467	0,3%	2.693	0,1%
Masculino	2.354.450	451.076	19,2%	1.056.956	44,9%	86.606	3,7%	749.272	31,8%	7.752	0,3%	2.788	0,1%
Total	4.591.806	882.130	19,2%	2.079.063	45,3%	165.628	3,6%	1.444.285	31,5%	15.219	0,3%	5.481	0,1%

Fonte: Censo Escolar (Inep 2021)

Quando analisadas individualmente, todas as escolas no âmbito deste projeto têm uma consistência semelhante de sub-representação de alunos negros/pardos. Ainda, valendo-nos dos dados da PNAD e INEP, elas apresentam uma grande diferença da proporção de alunos declarados como brancos ou pardos/negros.

Confrontado com outras escolas, o colégio *São Luís* tem uma porcentagem maior de alunos não declarados (11,58%), seguido por *Anchieta* (9,69%). A proporção de 3% de pessoas amarelas no *São Luís* e no *Sanfra* também é notável, assim como a proporção de 8% de alunos pardos em *Santo Inácio* e no *Loyola* (GRAF. 7).

Gráfico 7 – Composição % População Escolar FUND 2, ANEAS, por Raça/Cor e Unidade. (2018- 2021)

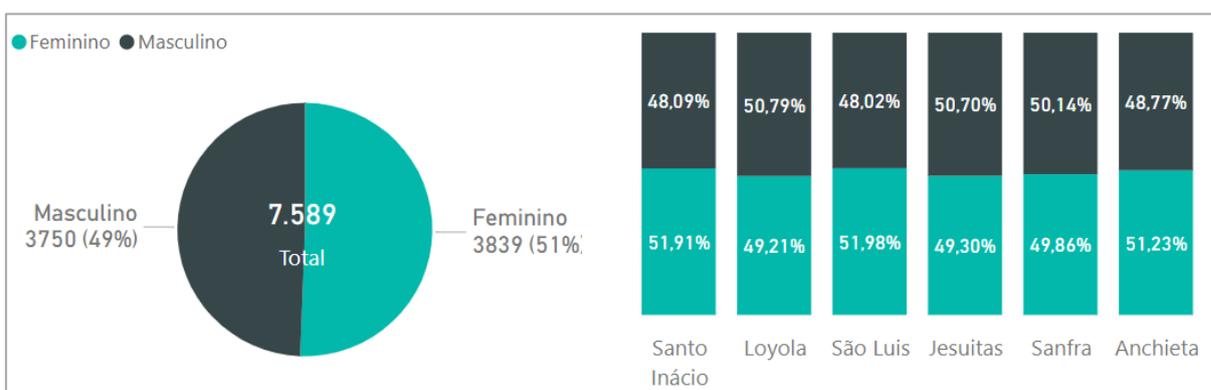


Fonte: Elaborado pelo autor.

O Gráfico 8 apresenta dados sobre a composição dos alunos por gênero nos colégios da ANEAS. No total, foram 7.589 alunos matriculados entre 2018 e 2021, considerando alunos distintos, ou seja, independente de quanto tempo eles estão no colégio.

Destes estudantes, há um equilíbrio geral entre os do sexo masculino e feminino: 49% e 51%, respectivamente. Além disso, não há diferença significativa entre estudantes masculinos e femininos em escolas específicas da ANEAS, exceto no que diz respeito às proporções de gênero nos colégios Loyola, Jesuítas e Sanfra, que têm uma porcentagem maior de estudantes masculinos (ver gráfico 8).

Gráfico 8 – % População Escolar FUND 2 ANEAS, por Gênero e Unidade. (2018- 2021)



Fonte: Elaborado pelo autor.

Ao contrastar a distribuição de gênero nos colégios da ANEAS com as unidades escolares da Região Sudeste, apresentada na Tabela 6, existe um equilíbrio, porém, conforme os dados do Inep (2021), a participação masculina é maior (51,3%). Esse dado difere daqueles consolidados nos colégios jesuítas, com participação feminina maior (51%). Examinando de forma individualizada, *Loyola*, *Sanfra* e *Jesuítas* aproximam-se da proporção de mais matriculados do gênero masculino em relação ao feminino, conforme apontado pelo Censo.

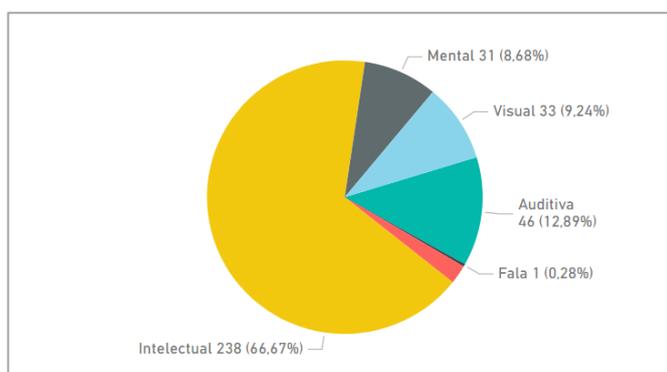
Tabela 6 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Sexo, Região Sudeste - 2021

Sexo				
Total	Feminino		Masculino	
4.591.806	2.237.356	48,7%	2.354.450	51,3%

Fonte: Censo Escolar (Inep 2021)

Quanto ao número de registros de deficiência, o Gráfico 9 apresenta a quantidade e o percentual de alunos matriculados entre 2018 e 2021 que registraram algum tipo de deficiência. No período, foram matriculados 315 alunos, com a maior incidência de estudantes com o tipo de deficiência intelectual (238), o que representa 66,67%; seguindo de auditiva (36); visual (33) e mental (31). Em todo o período, foi apontado apenas um aluno com o tipo de deficiência na fala.

Gráfico 9 – Quantidade alunos matriculados com Deficiência FUND2, ANEAS, por Tipo (2018-2021)

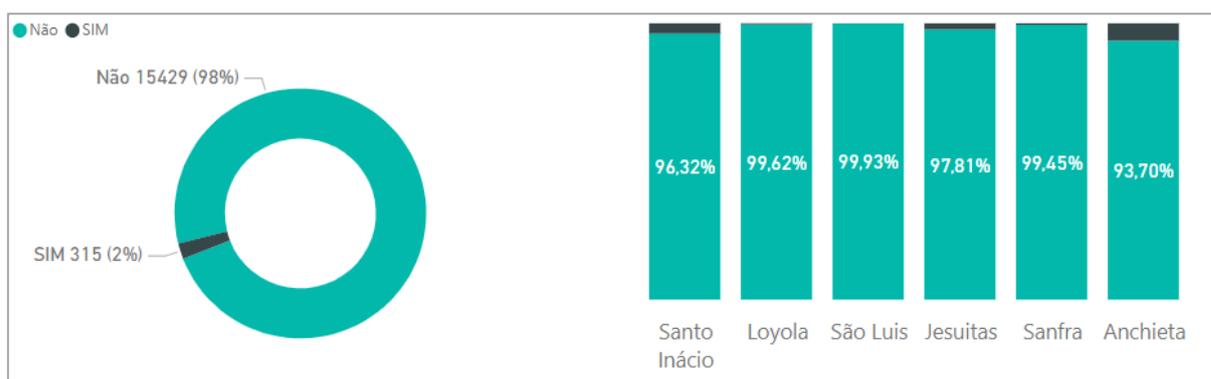


Fonte: Elaborado pelo autor.

O Gráfico 10 mostra o número de estudantes com algum tipo de deficiência. Em relação ao total de alunos distintos matriculados, o número absoluto de estudantes

com deficiência é de 2,07% (157 alunos). O *Anchieta* (RJ) apresenta o maior percentual de inclusão de deficientes, com 43 alunos (0,57% do total e 6,62% em relação aos alunos da unidade) e, considerando os registros no sistema ERP, o *São Luís* (SP) é o que possui o menor percentual de inclusão, com um aluno no período analisado.

Gráfico 10 – % População Escolar com Deficiência FUND2, ANEAS, por Unidade e Total (2018-21)



Fonte: Elaborado pelo autor.

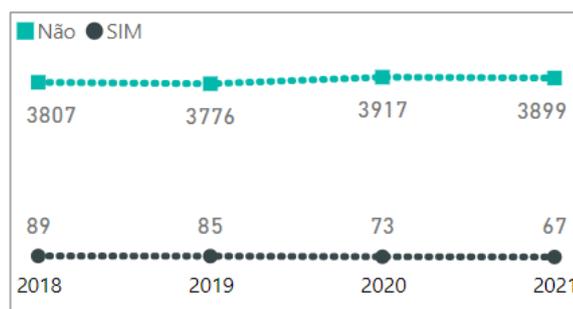
Ao correlacionar a análise dos colégios jesuítas estudados (2,07%), com os dados de 2021 do Censo, podemos evidenciar que na ANEAS a inclusão de deficientes é quase três vezes menor que a média das demais escolas na Região Sudeste, compreendendo também as públicas e rurais.

Examinando os dados do Censo da Educação Especial, em 2021, na Região Sudeste, nos anos finais do Ensino Fundamental, foram 141.937 matrículas, o que representa 6,3% do total de alunos matriculados (4.591.806). Inclui matrículas de alunos com algum tipo de deficiência, transtorno global do desenvolvimento ou altas habilidades/superdotação em turmas do 6º ao 9º ano.

O Gráfico 11 apresenta a série histórica de registro de alunos matriculados na ANEAS, em que “não” indica os alunos que não possuem registro de algum tipo de deficiência, e, “sim” alunos com registro de pelo menos um tipo de deficiência.

Mesmo com baixa inclusão de alunos com deficiência em relação ao total de matrículas, ela vem diminuindo a cada ano. Em 2018, eram 89 estudantes com algum tipo de deficiência *versus* 67, em 2021. Uma redução de 25%, seguindo um movimento contrário ao da linha de alunos matriculados sem deficiência, que teve um crescimento de 2% no mesmo período (3.807 em 2018 para 3.899 no ano de 2021).

Gráfico 11 – População Escolar com Registro de Deficiência FUND2, ANEAS, por Período Letivo (2018-2021)

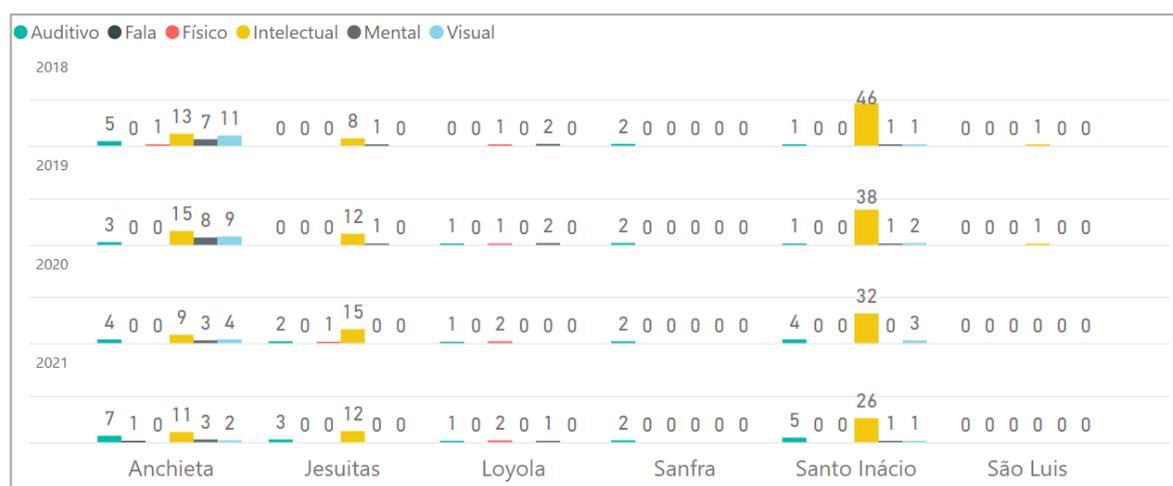


Fonte: Elaborado pelo autor.

O Gráfico 12 mostra a distribuição dos alunos com deficiência por tipo, unidade e ano letivo (2018-2021). Na série histórica, os colégios *São Luís*, *Sanfra* e *Loyola* tiveram pouquíssimos alunos com algum tipo de deficiência. Por exemplo, o *São Luís* não possui registros de deficiência em 2020 e 2021. Portanto, como este projeto é baseado em informações do ERP, cabe reforçar à gestão educacional avaliar periodicamente se esses registros do sistema refletem, realmente, a situação real da instituição.

Em 2021, o *Santo Inácio* registrou 33 alunos, sendo o maior número de alunos com algum tipo de deficiência, seguido por *Anchieta* (com 22), e *Jesuítas* (15 estudantes). Nessas escolas, a maior prevalência do tipo de deficiência foi a 'Intelectual', seguida pela "Auditiva".

Gráfico 12 – Número de Matrículas FUND2, ANEAS, por Tipo de Deficiência, Unidade e Período Letivo. (2018 – 2021)



Fonte: Elaborado pelo autor.

O Conselho Nacional de Educação (CNE) – Câmara de Educação Básica, através da Resolução Nº 3, de 3 de agosto de 2005, define normas nacionais para a ampliação do Ensino Fundamental para nove anos. No seu artigo 2º explicita as etapas do ensino básico e a idade ideal, aqui destacada a etapa do ensino fundamental:

Art.2º A organização do Ensino Fundamental de 9 (nove) anos e da Educação Infantil adotará a seguinte nomenclatura: Educação Infantil -Creche: Faixa etária - até 3 anos de idade - Pré-escola: Faixa etária -4 e 5 anos de idade. **Ensino Fundamental** de nove anos - até 14 anos de idade. Anos iniciais - Faixa etária de 6 a 10 anos de idade - duração 5 anos. Anos finais - Faixa etária de 11 a 14 anos de idade - duração 4 anos.

A pesquisa realizada revelou que os alunos das escolas estudadas se enquadram em duas faixas etárias principais: 11 a 14 anos e entre 15 e 17 anos. Das 15.744 matrículas na ANEAS, na série histórica analisada, 87,5% dos alunos tinham entre 11 e 14 anos nos respectivos períodos letivos. Na faixa de 15 a 17 anos, há 12,5% dos estudantes; e menos de 1% estão nas faixas de até 10 anos e de 18 a 19 anos (Tabela 7).

Tabela 7 – Faixa Etária Matrículas no Ensino Fundamental Anos Finais, ANEAS, por unidade (2018-2021).

	Faixa etária								
	Total	Até 10 anos		11 a 14 anos		15 a 17 anos		18 a 19 anos	
Anchieta	1.365	-	0,0%	1.136	83,2%	228	16,7%	1	0,1%
Jesuítas	2.373	-	0,0%	2.196	92,5%	177	7,5%	-	0,0%
Loyola	3.727	-	0,0%	3.077	82,6%	650	17,4%	-	0,0%
Sanfra	1.449	-	0,0%	1.353	93,4%	95	6,6%	1	0,1%
Santo Inácio	4.153	3	0,1%	3.556	85,6%	594	14,3%	-	0,0%
São Luís	2.677	-	0,0%	2.456	91,7%	221	8,3%	-	0,0%
TOTAL	15.744	3	0,0%	13.774	87,5%	1.965	12,5%	2	0,0%

Fonte: Elaborado pelo autor

Comparando as unidades, o Loyola, Anchieta e Santo Inácio possuem o maior percentual de alunos fora da faixa etária ideal, acima da média de 12,5%, em contrapartida os colégios Sanfra, Jesuítas e São Luís o menor percentual de alunos fora da faixa de idade ideal.

De acordo com o Censo 2021 (Inep), dos estudantes regulares do ensino fundamental, 88,4% daqueles que estão na Região Sudeste estavam matriculados na

faixa etária de 11 a 14 anos, a porcentagem mais alta entre todas as outras faixas etárias.

A faixa etária de 15 a 17 anos foi a de segunda maior concentração, com 10,3%. Assim, em relação às outras escolas da Região Sudeste, o *Jesuítas*, o *Sanfra* e *São Luís* têm percentuais mais altos na faixa etária ideal (11-14 anos), do que a média das escolas da região, sugerindo um menor número de alunos com distorção idade/série, que será apresentada abaixo (Tabela 8).

Tabela 8 - Número de Matrículas nos Anos Finais do Ensino Fundamental Regular, por Faixa Etária, Região Sudeste – 2021

Faixa etária								
Total	Até 10 anos		11 a 14 anos		15 a 17 anos		18 a 19 anos	
4.586.080	46.588	1,0%	4.055.202	88,4%	470.538	10,3%	13.752	0,3%

Fonte: Censo Escolar (Inep 2021)

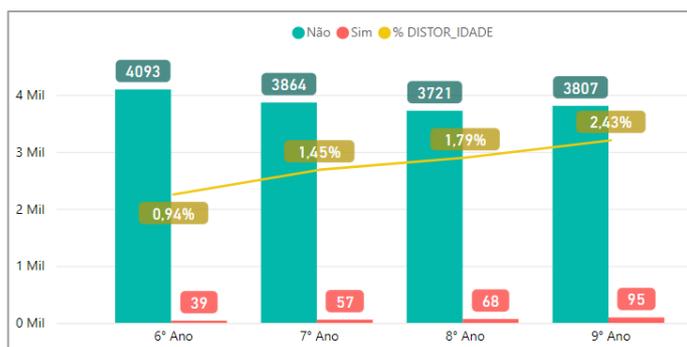
O Censo Escolar 2021 descreve que há três grandes momentos em que os índices de distorção idade-série⁸ são mais altos: o 3º ano o 6º ano do ensino fundamental, e o 1º ano do ensino médio. Esses três pontos críticos coincidem com as etapas de transição no percurso escolar de crianças e adolescentes: o final do ciclo de alfabetização, a mudança da sala de aula unidocente⁹ para a multidocente, entre outros.

Ao analisar os dados da ANEAS, observamos que o 9º ano apresenta a maior porcentagem de distorção idade/série, com 2,43%. Em seguida, o 8º ano, com 1,79%. Outro aspecto é que o 6º ano, diferente dos dados do Censo, representa a porcentagem mais baixa, com 0,94%. Outra característica que se destaca na ANEAS é que esta distorção começa a aumentar a partir do 6º ano (Gráfico 13).

⁸ Distorção ou defasagem idade-série é quando a diferença entre a idade do aluno e a idade prevista para a série é de dois anos ou mais.

⁹ Silva e Krug (2008) esclarecem que o prefixo 'uni' é originado da palavra unir e 'docência' corresponde ao ato do professor exercer sua profissão, então, logicamente 'unidocência' é a união dos conteúdos de todas as disciplinas e ministradas por um único professor.

Gráfico 13 – Distorção Idade-Série FUND 2, ANEAS, por ano-série (2018 – 2021)

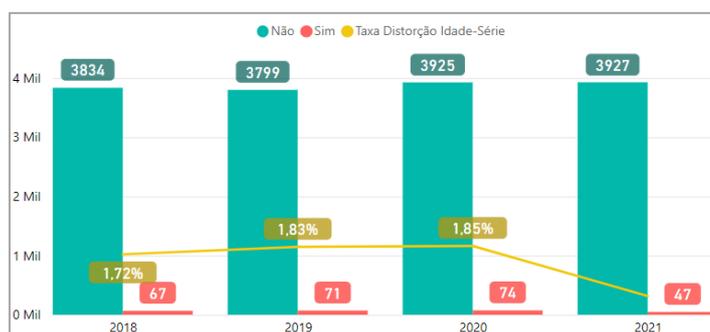


Fonte: Elaborado pelo autor

No Gráfico 14, a porcentagem de estudantes com distorção idade-série em 2021 caiu para 47 estudantes, reduzindo a taxa de distorção para 1,18%, a partir de uma média histórica de 1,8%.

A taxa de distorção idade-série na ANEAS é relativamente baixa, se comparada com taxa distorção idade-série dos colégios privados da região Sudeste e localização urbana. Segundo o Censo (Inep, 2021), em 2021, a taxa de distorção nesses colégios foi de 4,4%.

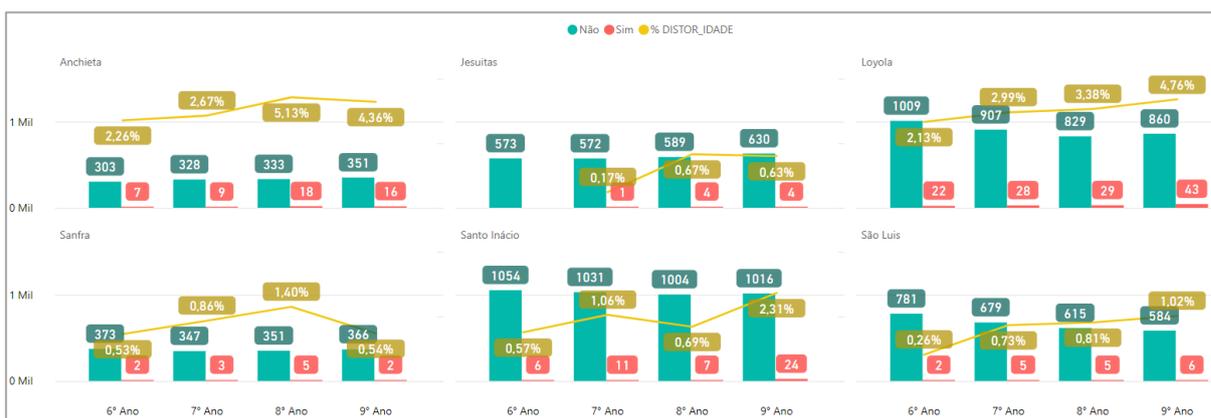
Gráfico 14 – Taxa distorção idade-série FUND2, ANEAS, por período letivo (2018-2021)



Fonte: Elaborado pelo autor

O Gráfico 15 apresenta a distorção idade/série por colégio. As distorções do *Anchieta* e do *Loyola* são as mais altas em relação à média dos colégios da ANEAS apresentadas no gráfico anterior. Os anos e os colégios com as maiores distorções foram no 8º e 9º ano, no colégio *Anchieta*, com 5,13% e 4,36% respectivamente; e da mesma forma, no 8º e 9º ano no *Loyola*, com 3,38% e 4,76%.

Gráfico 15 – Número e % de Matrículas nos Anos Finais do Ensino Fundamental Regular, ANEAS, por Distorção Idade-Série, Colégio e Ano (2018 – 2021)



Fonte: Elaborado pelo autor

Os alunos dos colégios *Jesuítas*, *Sanfra* e *São Luís* apresentaram a menor porcentagem de distorção nos quatro anos do ensino fundamental estudados. Entre os colégios, o *Jesuítas* registrou percentuais de distorção bem abaixo da média da ANEAS e, na série histórica (2018-2021), não computou nenhum aluno com distorção no 6º ano.

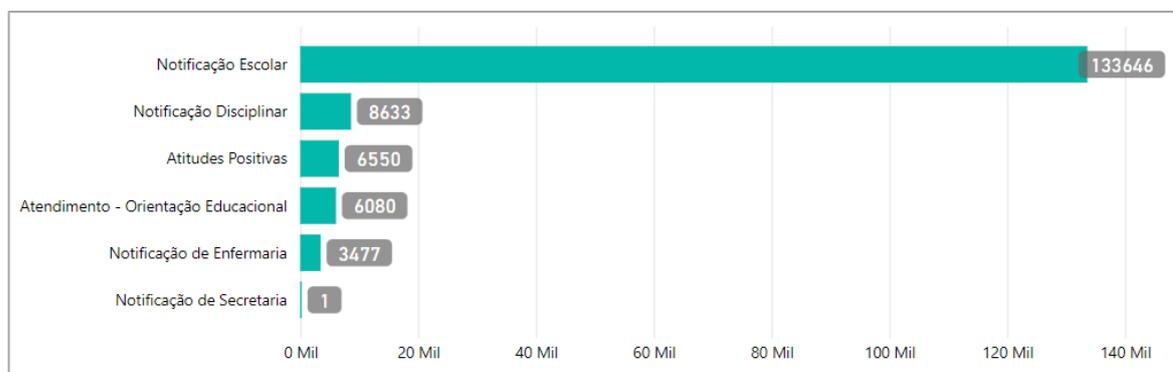
5.2.2 Dados Acadêmicos

O registro de ocorrências fornece às instituições educacionais um acompanhamento do desenvolvimento de atitudes e situações de seus alunos. Na ANEAS, eles são organizados em grupos, com os respectivos tipos de ocorrências, que são únicas e compartilhadas entre escolas. As ocorrências podem ser registradas pelo Corpo Docente, Coordenação Pedagógica, Secretária, Serviço de Orientação Educacional (SOE), Enfermaria, Biblioteca e Tesouraria.

O Gráfico 16 mostra que, nos períodos letivos de 2018 a 2021, um total de 158.387 ocorrências foram incluídas no sistema de gestão pelos colégios, distribuídas por grupos de tipos de ocorrência. A "notificação escolar", um grupo de tipos de ocorrências relacionadas a problemas de desempenho dos alunos em sala de aula, por exemplo, tarefa não realizada, falta de material, tarefa incompleta, entrada tardia na sala de aula, conversas entre outros, listou o maior volume de registros, com 133.646 ocorrências representando 84% do total.

As ocorrências do grupo "Notificação Disciplinar", associadas a comportamento inadequado ou infracional no ambiente escolar totalizaram 5% dos registros, com 8.633 ocorrências. Já os grupos de ocorrências "Atitudes Positivas" e "Orientação Educacional" registraram um percentual de 4%, com 6.550 e 6.080, registros respectivamente.

Gráfico 16 – Número de Ocorrências FUND2, ANEAS, por Grupo (2018 – 2021)



Fonte: Elaborado pelo autor.

A partir dos percentuais apurados, podemos evidenciar que a área acadêmica é responsável pelo maior volume de registros. Logo, vale a reflexão da gestão acadêmica se existe o engajamento das demais áreas da escola para registrar todas as ocorrências no sistema administrativo, uma vez que esses dados são importantes para a análise de perda de alunos.

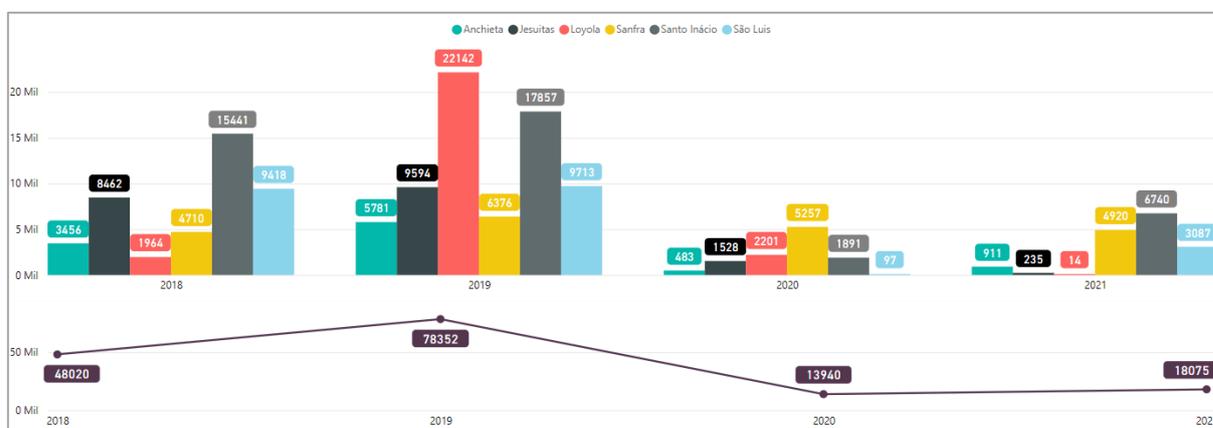
O Gráfico 17 evidencia que houve uma forte diminuição nos registros de ocorrência no sistema de gerenciamento durante o período pandêmico da covid-19, em comparação com períodos anteriores. Por exemplo, uma queda de 82,2% nos registros em 2020 (13.940), que foi o início da pandemia no Brasil, em comparação com 2019, que teve o maior volume de registros da série histórica, com 78.352 ocorrências registradas.

Podemos entender que essa redução possa estar associada ao uso de atividades não presenciais, aulas remotas e fora do ambiente escolar, que constituíram uma forma utilizada pelas escolas na tentativa de minimizar perdas causadas por restrições e recomendações de isolamento social.

Em 2021, com o Brasil ainda atingido pela covid-19, esperava-se que os volumes de registro caíssem. No entanto, houve um pequeno aumento, 18.075 em relação aos 17.904 do ano anterior, mas ainda muito inferior aos anos sem pandemia.

A questão que se apresenta é se, com o retorno do comparecimento em sala de aula, como resultado da vacinação e da flexibilização das restrições, os volumes de registro serão retomados nos próximos anos.

Gráfico 17 - Número de Ocorrências FUND 2, ANEAS, por Unidade e Período Letivo (2018 – 2021)



Fonte: Elaborado pelo autor.

Explorando os dados por unidade do Gráfico 17, o *Sanfra* obteve a menor variação no volume de registros na série analisada, sendo distinto dos demais colégios, que seguiram o comportamento de redução súbita descrita anteriormente. Outro ponto relevante é o baixíssimo número de ocorrência no *Loyola* em 2021, com apenas 14 registros.

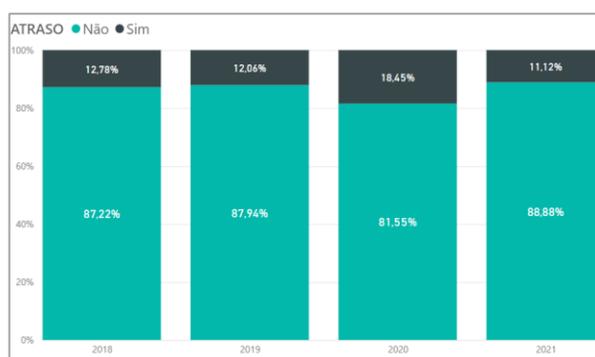
5.2.3 Dados Financeiros

Nos colégios jesuítas, para realizar a matrícula, o aluno deve estar adimplente, em outros termos, sem pendências financeiras com a instituição. Logo, alunos matriculados no ano corrente estão em dia ou renegociaram suas dívidas do ano anterior. Dessa forma, este estudo se valeu, para análise do relacionamento financeiro, dos pagamentos de mensalidades em atraso.

O Gráfico 18 apresenta o percentual de mensalidades que foram pagas com atraso acima de 30 dias, por período letivo. Na série histórica, observa-se o maior valor percentual de mensalidades pagas em atraso em 2020, com 18,45%. Importante lembrar que, naquele mesmo ano, foi confirmado o primeiro caso de pessoa contaminada com o coronavírus no Brasil. Posteriormente, teve início da adoção das medidas de mitigação, como o isolamento vertical ou horizontal.

Em 2021, como dito anteriormente, com níveis de contaminação da covid-19 ainda altos, foi registrado o menor percentual de pagamentos em atraso da série histórica (11,12%). Esse dado sinaliza uma eventual acomodação ou adaptação das famílias em relação aos impactos financeiros decorrentes das medidas de contenção adotadas pelos governos.

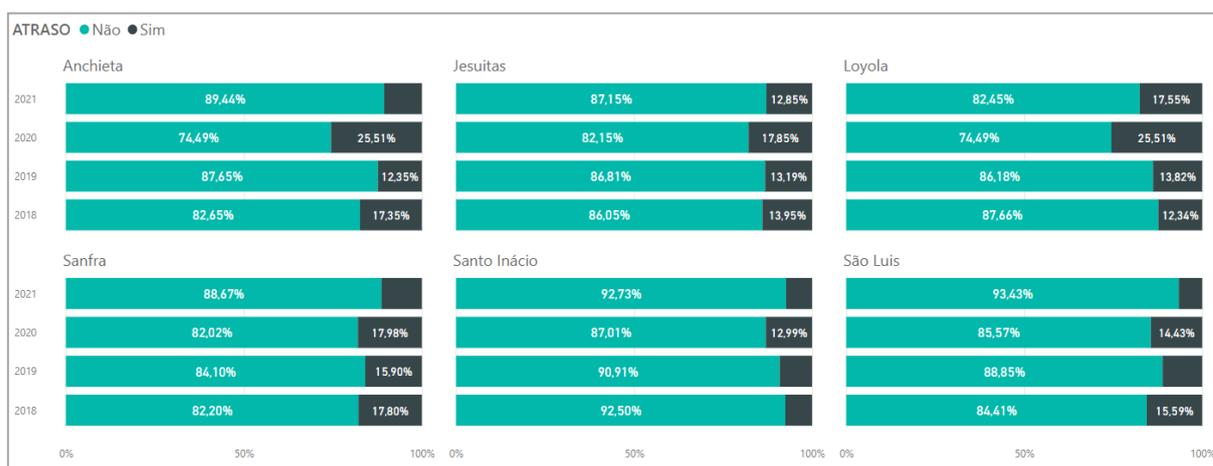
Gráfico 18 – % de Mensalidades pagos com ou sem atraso FUND 2, ANEAS, por Período Letivo (2018 – 2021)



Fonte: Elaborado pelo autor.

O gráfico 19 detalha as informações sobre os atrasos por unidade. Como apresentado na análise anterior, com exceção do *São Luís*, todas os colégios tiveram a maior porcentagem de pagamentos em atraso em 2020. Ao mesmo tempo, *Loyola* e *Anchieta* registraram a maior porcentagem de pagamentos em atraso em toda a série analisada, ambos com 25,51%.

Gráfico 19 – Porcentagem de Contratos pagos com ou sem atraso FUND2, ANEAS, por Período Letivo e Colégio (2018 – 2021)



Fonte: Elaborado pelo autor.

Em relação aos outros colégios, *Santo Inácio* sofreu o menor impacto nos atrasos de pagamento em 2020, com 12,99%. Nos demais anos de análise, apresentou uma média superior a 90% de pagamento, com menos de 30 dias do vencimento, sendo o melhor desempenho entre os coirmãos na série histórica.

5.2.4 Movimentação Acadêmica

Nesta seção, serão apresentadas as movimentações acadêmicas nos colégios da ANEAS no final de cada período escolar, ou seja, perda, retenção e entrada de novos alunos. Posteriormente, será feita a análise desses movimentos, correlacionando-os com os atributos definidos para entender a perda.

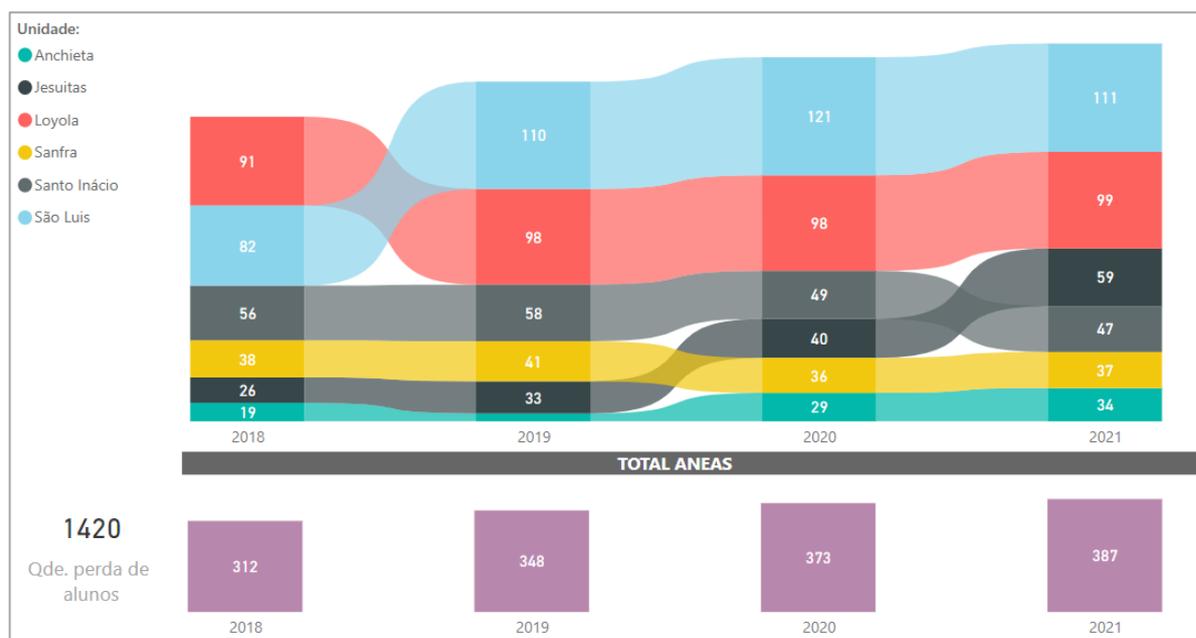
Para fins de cálculo de perda de alunos, considera-se o *status* de "Situação de Matrícula". Os valores absolutos de perda são obtidos pela soma das inscrições com o *status* de "situação de matrícula" de perda no respectivo período letivo, conforme detalhado na seção 5.1 deste projeto. A taxa da perda é o resultado da expressão: Soma dos alunos inscritos com *status* de "situação de matrícula" de perda / Total de alunos inscritos com *status* de matrícula de "perda e retenção", por período letivo.

O Gráfico 20 mostra a composição da perda de 1.420 alunos nos colégios da ANEAS, entre 2018 e 2021. Nele, observa-se um pequeno, mas constante crescimento na perda de alunos durante os períodos, com as maiores quantidades ocorrendo em 2020-2021 (373 e 387 alunos).

Analisando os resultados dos colégios individualmente, em números absolutos, o *São Luís* e o *Loyola* aparecem com a maior fuga nos últimos três anos. Por exemplo, em 2021, estes colégios perderam 111 e 99 alunos, respectivamente, no ensino fundamental anos finais. O colégio *Anchieta* apresentou o menor número absoluto em todos os períodos letivos, apesar de um crescimento nos anos pandêmicos, com 29 e 34 perdas sucessivamente.

A estabilidade na perda de alunos no *Loyola* – que permaneceu perto de 98 alunos nos últimos três anos – e a melhora no desempenho de *Santo Inácio*, que diminuiu sua perda em 2020 e 2021, para 49 e 47 alunos respectivamente, são dois aspectos a serem observados.

Gráfico 20 – Total de perda de alunos FUND 2, ANEAS, por colégio e período letivo. (2018-2021)



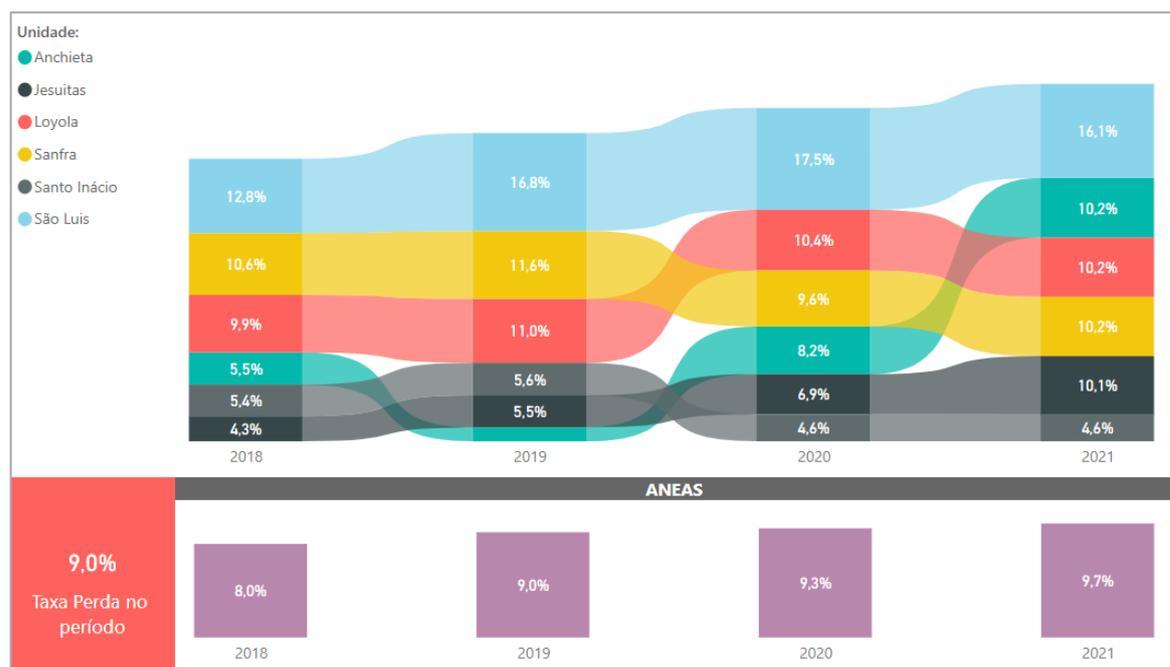
Fonte: Elaborado pelo autor

O Gráfico 21 mostra que a taxa de perdas na ANEAS – que foi de 9% na média – teve um crescimento constante desde o primeiro período acadêmico. Em 2020, acima da média, chegou a 9,3% e atingiu o nível mais alto em 2021, 9,7%.

Nos quatro períodos letivos, *São Luís*, *Sanfra* e *Loyola* tiveram porcentagens de perdas maiores do que as respectivas médias da ANEAS. Em 2021, com exceção de *Santo Inácio*, todas as escolas também tiveram perdas acima da média para aquele ano (9,7%), tendo o *São Luís* o maior percentual (16,1%). Na direção oposta, nesse mesmo ano, *Santo Inácio* registrou a menor porcentagem de perdas, com 4,6%, enquanto *Jesuítas*, *Sanfra*, *Loyola* e *Anchieta* obtiveram taxas próximas a 10,2%.

O crescimento na taxa de perda no *Anchieta* e no *Jesuítas*, no período 2018-2021, também foi digna de nota. Em 2019, a taxa no *Anchieta* foi de 2,2% – a mais baixa na série histórica. Em 2020 e 2021, esse índice subiu para 8,2% e 10,2%, respectivamente, uma evolução de 363,6%, em comparação com 2019, tornando-se a segunda maior taxa entre os colégios no ano de 2021. Já no colégio *Jesuítas*, em 2018, sua taxa era de 4,3% – a menor daquele ano – e, após vários crescimentos, atingiu 10,1% em 2021, uma variação de 134%.

Gráfico 21 – Taxa de Perda de alunos FUND2, ANEAS, por colégio e período letivo (2018-2021)



Fonte: Elaborado pelo autor.

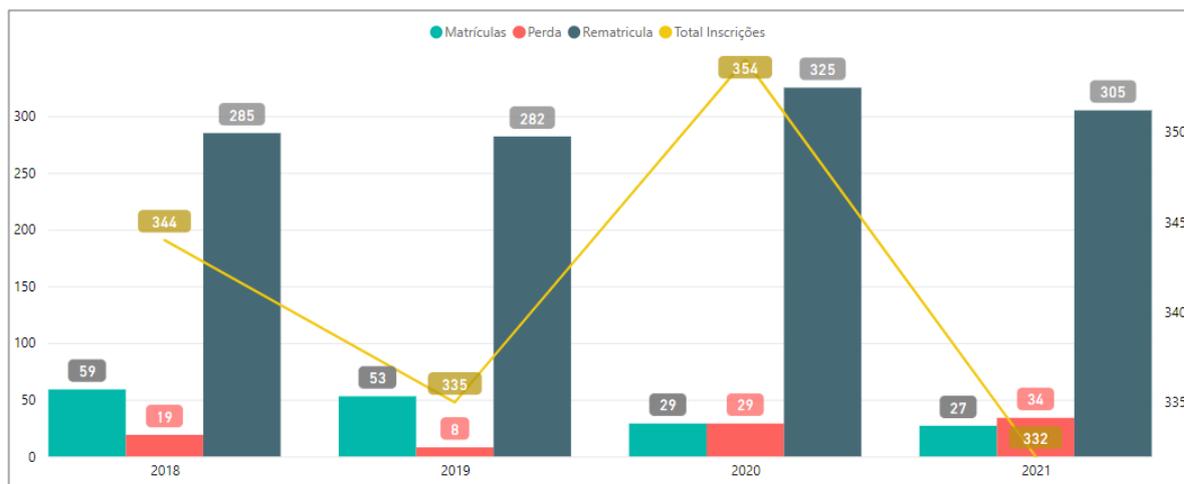
Sobre esses resultados, vale ressaltar o objetivo geral deste estudo, que é analisar se o sistema de detecção precoce (EDS), a ser proposto e aplicado no contexto do ensino fundamental, anos finais dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS), oferece informações que subsidiem a detecção precoce da perda de alunos. Quer dizer, este projeto não se aprofunda na compreensão do contexto e fatores que geram essas taxas de perda.

Nesse contexto, cabe destacar que a escola, como uma instituição que interage e serve à comunidade, enfrenta múltiplos desafios. Logo, é importante que a administração educacional das escolas entenda essas perdas, dentro do seu contexto. Por exemplo, o *São Luís*, em 2020, mudou de sede e proposta pedagógica, para a educação em tempo integral, e eventualmente o aumento das perdas pode ser o resultado do movimento estratégico de reposicionamento.

No entanto, apesar de não ser o objetivo deste projeto, através dos artefatos desenvolvidos para a compreensão dos dados empíricos relacionados à perda, alguns conhecimentos permanecem para estudos futuros. Como tal, o Gráfico 22, que detalha o resultado do colégio *Anchieta*, em 2019, ocorreu a sua menor taxa de perda, com a saída de oito alunos, menor quantidade da série histórica, além da entrada de 53 novos alunos (matrícula), a segunda melhor entrada. Contudo, em 2021, a sua

maior taxa de perda foi com 34 saídas de alunos e 27 matrículas, a menor da série histórica.

Gráfico 22 – Movimentação Acadêmica, Anchieta, por período letivo (2018-2021)



Fonte: Elaborado pelo autor

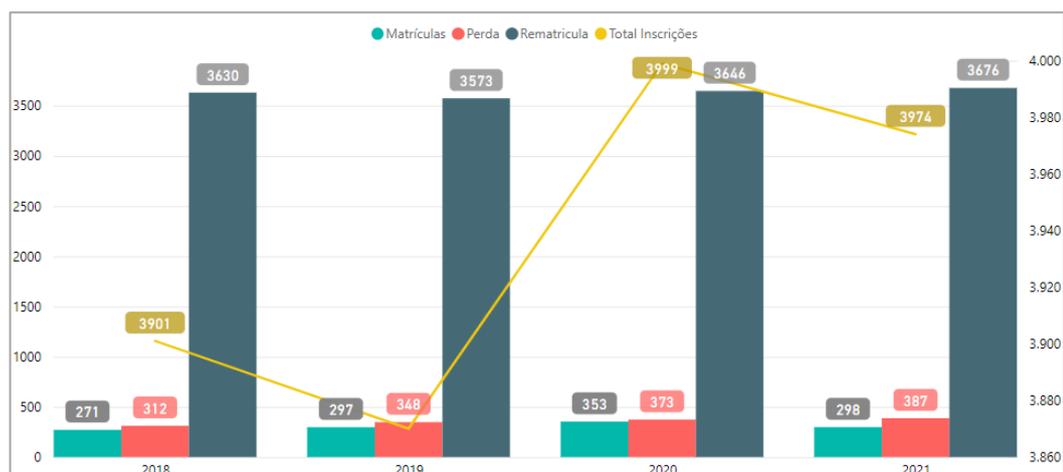
Em 2021, observa-se que houve uma diminuição nas rematrículas do *Anchieta*. Isso significa menor transição interna para o ensino fundamental. Esta variável associada à saída de alunos elevou a taxa de perda.

Assim sendo, os resultados do estudo atual indicam que, obviamente, a perda de alunos contribuiu para aumento da taxa de perdas. Entretanto, outras variáveis também podem interferir na taxa, como a transição interna de alunos: entrada de alunos do ensino fundamental nos primeiros anos para os anos finais (5º para 6º ano) e saída para o ensino médio (9º ano para a 1ª série do ensino médio). Essa movimentação pode contribuir para o aumento ou diminuição do número de rematrículas, afetando o cálculo da taxa de perda.

Ao aplicar essa mesma análise, no contexto ANEAS, isso pode justificar, mesmo com o aumento da taxa de perda, o aumento da população educacional no período pandêmico da covid-19, em comparação ao período anterior, como descrito na seção “Dados demográficos”.

O Gráfico 23 apresenta o crescimento das perdas. No entanto, da mesma forma, o número de entrada de novos estudantes externos (matrícula) e rematrículas – os quais, somados, são representados pela linha "população educacional" – foi maior em 2020 e 2021.

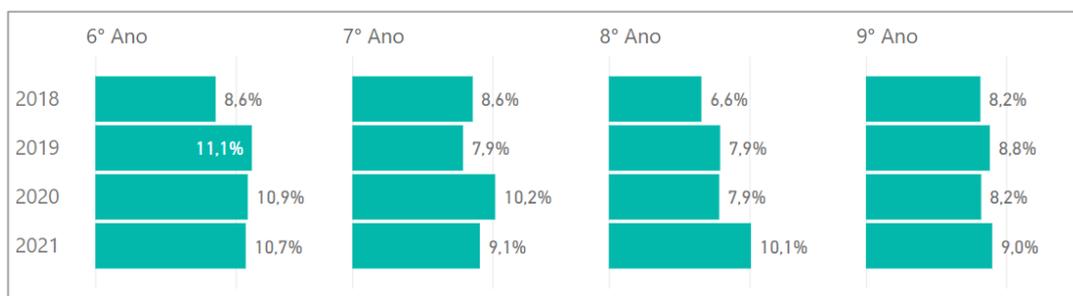
Gráfico 23 – Movimentação Acadêmica FUND2, ANEAS, por período letivo (2018-2021)



Fonte: Elaborado pelo autor

O Gráfico 24 retoma a análise da taxa de perda de alunos na ANEAS, apresentando os valores por ano-série e período letivo. O 6º ano obteve as maiores taxas de perda de alunos, seguido pelo 7º ano. Em 2021, ano que registrou a maior taxa de perdas (9,7%), o 6º e o 8º anos foram os que mais contribuíram, com 10,7% e 10,1%, respectivamente.

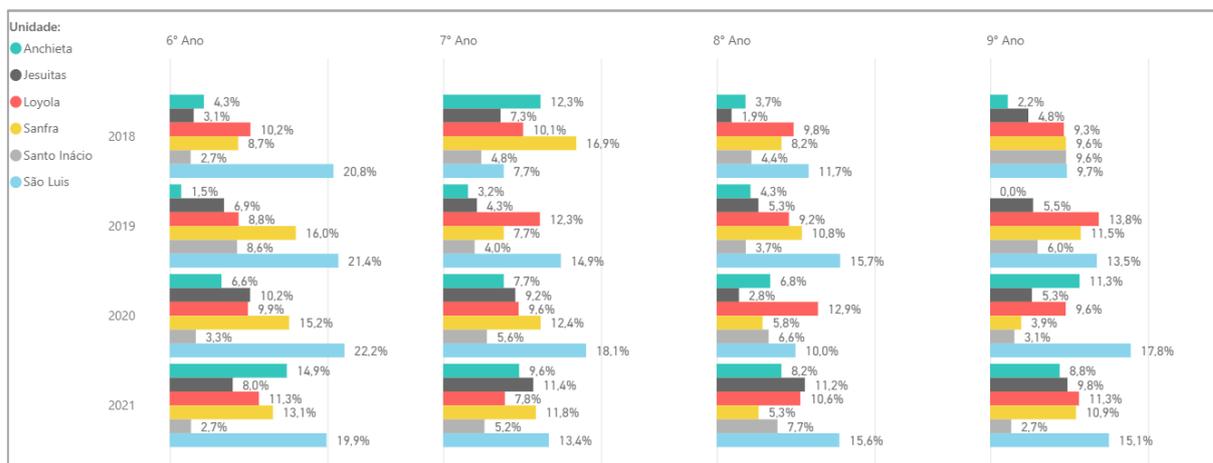
Gráfico 24 – Taxa de Perda de alunos FUND2, ANEAS por ano/série e período letivo (2018-2021)



Fonte: Elaborado pelo autor

Com relação à distribuição das perdas dos colégios entre os ano-série e períodos letivos, é possível ver uma clara diferença (conforme gráfico 25). Para algumas escolas, a taxa de perdas para um determinado ano-série é maior do que para outras. Por exemplo, em 2018, no *São Luís*, o 7º ano possuía taxa de perdas de 7,7%, enquanto em 2020 era de 18,1%.

Gráfico 25 – Taxa de Perda de Alunos por colégio, ano/série e período letivo. (2018-2021)



Fonte: Elaborado pelo autor

Esses dados indicam que cada fator pode ter diferentes graus de influência na retenção dos alunos, o que justifica a importância de a gestão escolar dos colégios jesuítas compreender a realidade de cada instituição, adotando instrumentos e processos de monitoramento específicos para cada uma.

Para compreender a relevância dos principais atributos definidos, as próximas tabelas buscam relacioná-los com as movimentações acadêmicas do atributo estudado: perda ou retenção.

A Tabela 9 mostra que, em relação ao gênero, a perda feminina foi ligeiramente maior (9,3%) do que a masculina, com 8,7%. A princípio, podemos inferir que no cenário estudado é uma variável de baixo impacto, já que se aproxima da taxa média de perda (9,0%).

Tabela 9 - Perda de alunos por Gênero, FUND 2, ANEAS (2018-21)

Gênero	Retido	%	Perda	%	Total
Feminino	7.251	90,7%	747	9,3%	7.998
Masculino	7.073	91,3%	673	8,7%	7.746
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

Na Tabela 10, a despeito de representar um percentual mais baixo (5,05%) em relação ao total da amostra, os alunos de raça/cor “Não declarada” têm um percentual de perda de alunos superior às demais raças, com 32,4%. Para os que se declaram

como cor “Branca”, que representam 87,61% do público total, a taxa de perda foi ligeiramente inferior à taxa média (9,0%), com 8%, ou seja, proporcionalmente alunos brancos saem menos do que a média de todas as raças.

Tabela 10 - Perda de alunos por Raça, FUND2, ANEAS (2018-21)

Raça/Cor	Retido	%	Perda	%	Total
Amarela	122	80,8%	29	19,2%	151
Branca	12.851	92,0%	1.117	8,0%	13.968
Indígena	5	100,0%	0	0,0%	5
Não declarado	416	67,6%	199	32,4%	615
Parda	842	92,1%	72	7,9%	914
Preta/Negra	88	96,7%	3	3,3%	91
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

No quesito deficiência, verifica-se que alunos que apresentam algum tipo de deficiência possuem taxa menor de perda (8,3%), em relação aos que não possuem (9,0%), conforme expresso na Tabela 11. Essa variável também se aproxima da taxa global (9,0%) e, por representar uma ocorrência muito pequena de alunos, não deve interferir no modelo preditivo.

Tabela 11 - Perda de alunos com deficiência, FUND2 ANEAS (2018-21)

Com deficiência	Retido	%	Perda	%	Total
Não	14.035	91,0%	1.394	9,0%	15.429
Sim	289	91,7%	26	8,3%	315
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

A Tabela 12 expressa as perdas de alunos em relação à sua idade. As maiores perdas proporcionais se concentram nos alunos de 17 e 16 anos, com taxas de 25% e 20%, respectivamente, mas com baixa representação absoluta, com a perda de apenas 20 estudantes. Alunos com 11 anos possuem um índice de perda de 11,8%, superior à média global, além de ser a idade com a quarta maior perda absoluta de alunos (223), o que indica uma maior tendência de perda nesta idade.

Tabela 12 - Perda de alunos por Idade, FUND2, ANEAS (2018-21)

Idade no período Letivo (anos)	Retido	%	Perda	%	Total
10	3	100,0%	0	0,0%	3
11	1.670	88,2%	223	11,8%	1.893
12	3.641	90,6%	379	9,4%	4.020
13	3.636	91,9%	319	8,1%	3.955
14	3.601	92,2%	305	7,8%	3.906
15	1.693	90,7%	174	9,3%	1.867
16	72	80,0%	18	20,0%	90
17	6	75,0%	2	25,0%	8
18	1	100,0%	0	0,0%	1
19	1	100,0%	0	0,0%	1
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

Segundo análise de Queiroz (2002), fundamentada em algumas entrevistas realizadas com os docentes, as causas para a evasão, são: a defasagem série/idade, brigas em sala de aula, bagunça, desrespeito e violência com os professores.

Como descrito na seção anterior, de acordo com a legislação brasileira, a faixa etária de escolarização obrigatória vai de 4 a 17 anos. Por lei, aos 4 anos, a criança deve ingressar na pré-escola; aos 6 anos, no ensino fundamental e, aos 15 anos, no ensino médio. No aspecto defasagem série-idade, ou seja, alunos que têm dois ou mais anos de atraso escolar a Tabela 13, a seguir, apresenta a idade adequada para o ensino fundamental.

Tabela 13 – Idade adequada Ensino Fundamental

Ano/Série	Idade Adequada
1º. ano	6
2º ano/1ª Série	7
3º ano/2ª Série	8
4º ano/3ª Série	9
5º ano/4ª Série	10
6º ano/5ª Série	11
7º ano/6ª Série	12
8º ano/7ª Série	13
9º ano/8ª Série	14

Fonte: Inep (2021).

Ao levar em consideração esse aspecto, nota-se que a distorção idade-série é uma variável a ser considerada, mas com número absoluto pouco representativo, com

48 perdas no período (conforme Tabela 14). A saída de estudantes nessas situações é quase o dobro daquela com a faixa etária apropriada para o ano.

Tabela 14 - Perda de alunos por “Distorção Idade Série”, Fund2, ANEAS (2018-21)

Distorção Idade Série	Retido	%	Perda	%	Total
Não	14.113	91,1%	1.372	8,9%	15.485
Sim	211	81,5%	48	18,5%	259
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

Na Tabela 15, a fim de determinar se o aluno recebeu a classificação ocorrência “sim”, foi necessário considerar se o aluno tinha pelo menos um registro no período escolar avaliado em um dos tipos de ocorrências "notificação escolar", "notificação disciplinar" e "orientação educacional".

Ainda, a fim de equalizar os dados dos alunos veteranos, aqueles que estão na escola há mais de um ano, foi verificado também se eles receberam uma ocorrência no período letivo anterior, para considerar a ocorrência como "sim" ou "não". É importante observar que muitas saídas de alunos ocorrem no início do período letivo, quando ainda não há registros de novas ocorrências. Ressalta-se que esse fato poderia distorcer a análise.

A maior taxa de perda de alunos é encontrada no grupo que não recebeu nenhuma ocorrência, com 16,3%. Os estudantes classificados como "sim" tiveram uma taxa de perda no período avaliado de 6,2%.

Tabela 15 – Perda de alunos por Ocorrência, Fund2, ANEAS (2018-21)

Ocorrência	Retido	%	Perda	%	Total
Não	3.713	83,7%	721	16,3%	4.434
Sim	10.611	93,8%	699	6,2%	11.310
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

Conforme Tito (1993), alunos nos primeiros anos da vivência escolar passam por um processo de adaptação na nova instituição, tornando-se estudantes “calouros”, classificados apenas com “matrícula” e, por isso, mais propensos a sair da escola.

Seguindo essa linha de pensamento, neste estudo, a Tabela 16 verifica o percentual de perda e retenção do aluno “calouro” em relação ao “veterano”. Embora

o maior volume de perda de alunos em número absoluto seja de “veteranos” (1.230), é possível afirmar que o maior índice percentual de perda está entre os estudantes “novatos” (15,6%), coerente com a análise de Tito (1993).

Tabela 16 - Perda de alunos por “Tipo de Matrícula”, Fund2, ANEAS (2018-21)

Por tipo de Matrícula	Retido	%	Perda	%	Total
Matrícula	1.029	84,4%	190	15,6%	1.219
Rematrícula	13.295	91,5%	1.230	8,5%	14.525
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

Ao analisar a perda de alunos associada aos pagamentos em atraso - aqueles realizados em data posterior a 30 dias do vencimento –, os dados registrados mostram que a porcentagem foi menor do que aqueles que pagaram até 30 dias (conforme Tabela 17). Isso implica que o atraso no pagamento das mensalidades escolares não é um fator determinante para a perda de alunos nos colégios da ANEAS. Portanto, considera-se que análises futuras devem ser realizadas para entender melhor a relação entre atraso de mensalidade e perda de alunos.

Tabela 17 - Perda de alunos por “Pagamento em atraso”, Fund2, ANEAS (2018-21)

Pagamento Atraso	Retido	%	Perda	%	Total
Não	12.236	90,7%	1.261	9,3%	13.497
Sim	2.088	92,9%	159	7,1%	2.247
Total	14.324	91,0%	1.420	9,0%	15.744

Fonte: Elaborada pelo autor.

A análise exploratória permitiu compreender o quadro geral da perda de alunos na ANEAS, relacionando-o com alguns atributos identificados em estudos semelhantes. Além disso, na análise individualizada dos dados dos colégios estudados – apesar de inicialmente apresentarem características semelhantes, como o público atendido, modelos pedagógicos e estar na mesma região do Brasil –, notou-se que eles apresentam conjuntos de dados e taxas de perdas de alunos diferentes. Isso sugere, portanto, que cada colégio tem seus próprios desafios.

5.2.5 Principais achados da análise exploratória

O Quadro 7 fornece uma visão consolidada das principais descobertas descritas nas seções anteriores, que apoiaram o desenvolvimento da pesquisa ou que deixaram caminhos para pesquisas posteriores.

Quadro 7 – Resumo Achados Análise Exploratória

Tema	Item	Achado
Dados demográficos	1	No período pandêmico da covid-19 (2020 e 2021), a ANEAS registrou a maior população escolar no FUNDEB em comparação com o período pré-pandêmico, sugerindo que, em certa medida, as unidades absorveram os impactos causados pelas políticas públicas para conter a propagação do contágio.
	2	Maior concentração de alunos matriculados no 6º e 7º ano.
	3	Os dados da ANEAS mostram uma alta representação dos alunos declarados como brancos e a sub-representação de pardos/preto quando comparados aos dados do INEP para escolas da mesma região.
	4	O percentual de inclusão de estudantes com deficiência na ANEAS é inferior à média das escolas da região Sudeste, de acordo com dados do INEP. Na série histórica, 2021 mostrou o menor percentual de inclusão.
	5	Redução da taxa de distorção idade-série em 2021 na ANEAS, para 1,18%. Conforme dados do Censo, no mesmo ano nos colégios privados da região Sudeste o percentual de distorção no Fundeb foi de 4,4%
Dados Acadêmicos	6	Redução significativa nos registros de ocorrência de estudantes nos sistemas de gestão da ANEAS nos anos pandêmicos, 2020 e 2021.
Dados financeiros	7	Em 2021, o menor índice de atraso de pagamento na ANEAS dos quatro períodos letivos avaliados, mesmo durante o período pandêmico.
Movimentação acadêmica	8	Crescimento do número absoluto de alunos perdidos a partir do início da série histórica, 2018.
	9	A taxa de perdas de alunos desde o início da série histórica (2018) na ANEAS está em ascensão. Ao analisar cada colégio individualmente, alguns tiveram uma diminuição nas perdas.
	10	Maior taxa de perda de alunos para alunos com raça/cor não declarada 32,5%, contra a média de 9%. Maior taxa de perda de alunos com idades 17, 16 e 11 anos. Entretanto, alunos com 16 e 17 anos com número absoluto baixo.
	11	Maior taxa de perda de alunos que não receberam ocorrência e de novatos

Fonte: Elaborado pelo autor.

6 EDS MODELO PREDITIVO

O capítulo anterior apresentou o entendimento dos dados históricos dos colégios da ANEAS, tendo como fio condutor a perda de alunos. A partir do objetivo geral deste estudo, foi aplicado o *Early Detection Systems (EDS)*, que se refere a um modelo preditivo para prever o resultado futuro de perdas de alunos baseado em eventos passados.

Neste estudo, ele se desenvolveu a partir dos dados consolidados dos colégios. Portanto, foi construído um EDS “genérico” ANEAS, posteriormente validado através da aplicação do modelo ao conjunto de dados consolidados e únicos de cada colégio.

O processo de criação do EDS parte dos conceitos de aprendizagem de máquina, um tipo de Inteligência Artificial que permite aos computadores aprenderem sem serem explicitamente programados. Segundo Michalski *et al.* (2013), o aprendizado de máquina é um conceito relacionado à área de Inteligência Artificial (IA), cujo objetivo é desenvolver métodos computacionais capazes de tomar decisões com base em conhecimentos já existentes.

É um subconjunto de um campo de estudo maior, conhecido como mineração de dados, que se preocupa com a extração de padrões de grandes conjuntos de dados. Para Williams (2011), a mineração de dados é a arte e a ciência inteligente de dados, cujo objetivo é descobrir informações significativas e conhecimento a partir dos dados.

O modelo preditivo foi elaborado utilizando o *software Weka*, que é um sistema de aprendizagem de máquinas desenvolvido pela Universidade de Waikato, na Nova Zelândia. O *software* é gratuito, de código aberto, e pode ser usado para treinar e testar algoritmos para tarefas de mineração de dados. Tem sido amplamente utilizado em ambientes acadêmicos, industriais e governamentais, devido a sua flexibilidade, à facilidade de uso, ao alto desempenho e à escalabilidade. A plataforma oferece algoritmos de mineração de dados e aprendizagem de máquinas que foram utilizados para descobrir os padrões a partir dos dados coletados no projeto, consolidado com a definição do EDS.

Nas próximas seções, será apresentado o percurso para viabilizar o EDS, utilizando técnicas de aprendizagem de máquinas, com o objetivo de analisar, ao final, se o sistema de detecção precoce (EDS), a ser proposto e aplicado no contexto do

ensino fundamental anos finais dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS), oferece informações que subsidiem a detecção precoce da perda de alunos.

6.1 ORGANIZAÇÃO E MODELAGEM DOS DADOS

No processo de modelagem preditiva, o primeiro passo é a criação dos bancos de dados de treinamento e testes. Neste caso, utilizamos dados de 2018 para treinar e testar nosso modelo de aprendizagem de máquinas.

Os atributos descritos na seção 5.1 (Tabela 6, p.74), foram considerados para a construção da base de dados. Eles foram submetidos a um processo de preparação, no qual foram discretizados e transformados em formas mais específicas para facilitar a identificação de padrões que limitam seus possíveis estados.

Não houve a necessidade de limpeza de dados e verificações de validade, como a remoção de registros duplicados ou a correção de entradas errôneas, pois elas foram feitas na análise exploratória. Entretanto, é importante observar que esse processo de saneamento de dados deve acontecer antes de passar para próxima etapa, do levantamento do ganho da informação, em que todos os dados são submetidos ao *software Weka*.

O ganho de informação é um método para determinar a importância dos atributos na mineração de dados. Ele compara a quantidade de informação obtida ao incluir cada atributo no modelo com a quantidade de informação perdida ao excluí-lo. Os atributos que obtêm maior valor são aqueles que apresentam um maior ganho de informação, conseqüentemente sendo os primeiros critérios de divisão para a árvore de decisão (WITTEN; FRANK, 2005). Quer dizer, se um atributo tem uma alta pontuação de ganho de informação, então, incluí-lo no modelo fornecerá mais informações úteis do que excluí-lo.

Na análise exploratória, observamos que alguns atributos, quando selecionados, apresentavam diferentes variações nas taxas de perdas entre os colégios. Dito de outro modo, um atributo selecionado em um colégio apresentava uma taxa de perda diferente em relação a outro colégio. Como mostrado no Quadro 8, essa informação foi considerada na organização dos conjuntos de dados, permitindo uma melhor compreensão do ganho de informação.

Cabe reforçar que os dados da base treinamento foram organizados em sete conjuntos de dados: o primeiro denominado ANEAS, com os dados consolidados dos colégios, e os demais por colégio, com seus respectivos dados. O conjunto de dados da ANEAS foi o condutor para a construção do EDS e os outros conjuntos permitiram a análise de cenários por unidade.

Quadro 8 – Conjunto dados para treinamento e teste (2018)

Nome base de dados	Quantidade Instâncias	Atributos	Algoritmo ganho informação	Ano Selecionado
ANEAS	3.901			
Anchieta	344	(Inclui o atributo alvo)	InfoGainAttributeEval [Use full training set]	2018
Jesuitas	605			
Loyola	922			
Sanfra	358			
SantoInacio	1.029			
SaoLuis	643			

Fonte: Elaborado pelo autor.

Para análise do ganho de informação do atributo alvo perda, optou-se pelo algoritmo *InfoGainAttributeEval*, da biblioteca *Weka*, utilizado em outros estudos de mineração de dados. O Quadro 9 apresenta o resultado da ANEAS, com a posição e o ganho de informação de todos os atributos. A mesma tabela mostra a posição dos atributos nos cálculos por colégio, destacando em amarelo os cinco com o maior ganho de informação. O Apêndice B deste estudo contém o resumo dos cálculos realizados no *Weka* das respectivas bases de treinamento e teste.

Quadro 9 – Ranking ganho de informação atributo alvo perda, base treino (2018)

ANEAS			COLÉGIOS					
Posição	Atributo	Ganho de Informação	Anchieta	Jesuitas	Loyola	Sanfra	Santo Inácio	São Luis
			Posição					
1	FALT_MAT	0.0943614	1	2	2	3	5	1
2	FALT_PORT	0.0871152	12	1	3	1	1	3
3	N_OCORRENCIA	0.0397951	3	3	4	2	3	2
4	RACA_ALUNO	0.0332126	5	6	1	10	7	6
5	OCORRENCIA	0.0302444	2	4	5	4	6	4
6	QDE_PARCELAS	0.0060614	13	12	12	12	13	12
7	DISTORCAO_IDADE	0.0034779	9	10	7	8	2	8
8	IDADE_NOPLETIVO	0.0025208	11	13	13	13	4	13
9	POSSUI_DEFICIENCIA	0.001029	8	9	9	9	9	11
10	TIPO_MATRICULA	0.0009469	7	7	6	6	10	9
11	SERIE	0.0006588	4	5	10	5	8	5
12	ATRASO	0.0002361	10	8	8	11	11	7
13	SEXO_ALUNO	0.0000191	6	11	11	7	12	10

Fonte: Elaborado pelo autor

No conjunto de dados ANEAS, o atributo com maior ganho de informação é FALT_MAT (quantidade de faltas na disciplina Matemática), com 0,0943614, tornando-se a possível raiz da árvore de decisão, seguido por FALT_PORT (quantidade de faltas na disciplina Português), com 0,0871152. Dentre os atributos de maior relevância, inclui-se o N_OCORRENCIA (quantidade de ocorrências), RACA_ALUNO (raça/cor do aluno) e OCORRENCIA (independentemente da quantidade, se recebeu ocorrência no período).

O algoritmo confirmou que alguns atributos têm pouco ganho de conhecimento, como observado na análise exploratória. Por exemplo, SEXO_ALUNO, com 0,0000191. Na Tabela 11, identificamos que as taxas de perda entre os sexos estão próximas umas das outras, indicando que é uma variável de pouco ganho de informação sobre a saída de estudantes. Assim, os demais atributos com baixo ganho de informação, eventualmente serão descartados na definição do modelo preditivo.

Quando avaliamos os resultados dos colégios, os atributos com maior ganho de informação estão agrupados na mesma área. Entretanto, o atributo SERIE tem maior relevância para quatro colégios, fora da posição quando comparado com o *ranking* no cálculo da ANEAS (11^o). Ainda, para o *Santo Inácio*, DISTORCAO_IDADE e IDADE_PLETIVO têm maior relevância.

Podemos inferir que um modelo de previsão – considerando os atributos de maior ganho de conhecimento avaliado no contexto da ANEAS – pode, até certo ponto, atender à previsão de todos os colégios. Entretanto, observa-se que, para maior precisão, é necessário analisar o ganho de conhecimento por unidade, gerando modelos preditivos específicos.

Também identificamos possíveis impactos no ganho de informação impostos pelo cenário da covid-19. Assim, um novo banco de dados de teste e treinamento foi estruturado considerando informações de 2020, o ano do início da pandemia no Brasil. Com essa nova base de dados, foram realizados os mesmos testes que antes, apresentando os resultados dispostos no Quadro 10.

Nessa análise, comparou-se o posicionamento dos atributos no cálculo anterior com o posicionamento do novo cálculo. Com dados de 2020, foi possível observar que os atributos relacionados à ocorrência (N_OCORRENCIA e OCORRENCIA) apresentam um maior ganho de informação, enquanto aqueles relacionados à ausência (FALT_MAT e FALT_PORT) perdem ganho. Outro aspecto a ser destacado é a maior dispersão da posição dos atributos entre os colégios, em comparação com

o *ranking* da ANEAS. Por exemplo, o atributo *DISTORCAO_IDADE* para o *Anchieta* tem o maior ganho de informação, enquanto para o *Jesuítas* ele ocupa a 13ª posição.

Quadro 10 - *Ranking* ganho de informação atributo alvo PERDA, base treino (2020)

ANEAS			COLÉGIOS					
			Anchieta	Jesuítas	Loyola	Sanfra	Santo Inácio	São Luis
Posição	Atributo	Ganho de Informação	Posição					
1	N_OCORRENCIA	0.055253	2	2	2	1	1	2
2	OCORRENCIA	0.027122	8	3	3	2	7	3
3	FALTA_MAT	0.012136	13	1	5	13	13	13
4	RACA_ALUNO	0.010854	10	8	8	5	6	1
5	FALTA_PORT	0.010374	11	12	1	12	12	9
6	TIPO_MATRICULA	0.005598	7	9	4	8	9	8
7	QDE_PARCELAS	0.002748	12	11	6	10	10	10
8	ATRASO	0.00269	9	7	7	7	4	7
9	DISTORCAO_IDADE	0.002639	1	13	12	4	3	5
10	SERIE	0.001395	5	4	9	3	2	4
11	SEXO_ALUNO	0.000701	4	10	11	9	8	6
12	POSSUI_DEFICIENCIA	0.000641	6	6	10	6	5	11
13	IDADE_NOPLETIVO	0	3	5	13	11	11	12

Fonte: Elaborado pelo autor.

Os resultados dessas análises sugerem que os ganhos de informação obtidos por colégios ou períodos letivos são diferentes, o que pode indicar que outros fatores influenciam seu comportamento e devem ser considerados no desenvolvimento do modelo preditivo. Eventualmente, um modelo genérico pode não se adaptar completamente à realidade de cada escola ou período letivo.

6.2 TREINAMENTO E TESTES

Para a indução da árvore de decisão foi utilizado o algoritmo J48. Segundo o estudo realizado por Norouzi *et al.* (2016) e por Sathya e Rajesh (2015), comparando a porcentagem de classificações corretas entre os algoritmos de classificação, o J48 tem uma das melhores taxas de precisão para a descoberta do conhecimento.

O treinamento, o teste e a avaliação do modelo foram organizados considerando duas opções de teste. A primeira, dividindo o banco de dados em 70% para treinamento do modelo, e o restante para testes (*Percentage Split*). A segunda, dividindo os dados em conjuntos (10 partes), em que um conjunto é usado para treinamento e o outro para testes (*Cross-validation*).

Neste estudo, a primeira opção foi escolhida porque permite assegurar que os melhores resultados possíveis sejam obtidos a partir do conjunto de dados. A segunda

opção foi escolhida porque ajuda a entender se conseguimos uma boa generalização de nosso modelo, pois nos permitirá ver se resultados semelhantes são obtidos por diferentes modelos, treinados em diferentes subconjuntos de dados.

O modelo preditivo (EDS) foi construído a partir do conjunto de dados de 2018, com as informações consolidadas da ANEAS (3.901 instâncias). Em outras palavras, neste projeto, foi desenvolvido um EDS “genérico”, que será validado posteriormente, com dados de cada colégio. O Quadro 11 apresenta a síntese de como os testes para definição do EDS foram organizados.

Quadro 11 – Organização treinamento e teste do Modelo Preditivo.

Treinamento e teste	Algoritmo para classificação	Opções de testes	Base dados	Quantidade de instâncias	Observação
1	J48	Percentage split (70%)	2018 (PRÉ-COVID-19)	3901	Divisão randômica pelo Weka de 70% das instâncias para treino e 30% para teste.
2		Cross-validation (Folds 10)			Particionamento pelo WEKA em 10 conjuntos, onde um conjunto será utilizado para treino e outro para teste.

Fonte: Elaborado pelo autor.

A próxima seção descreve os resultados dessas análises, considerando a interpretação do desempenho nos testes do modelo preditivo. A análise abrangeu tanto o desempenho geral do modelo quanto as previsões individuais.

6.2.1 Treinamento e teste base (2018)

Na abordagem inicial no *Weka*, o algoritmo J48 gerou a árvore de decisão, baseada no conjunto de dados 2018, contendo os 14 atributos, independentemente do ganho de informação previamente avaliado. Tal ação resultou no modelo preditivo (EDS) treinado para testes.

Como aplicado nos estudos de Norouzi *et al.* (2016) e nos de Sathya e Rajesh (2015), a interpretação do desempenho do algoritmo J48 é analisada a partir da estatística de precisão de classificação correta e incorreta das instâncias, validando se o modelo proposto obteve um aprendizado supervisionado satisfatório.

Na análise de desempenho “Treinamento Teste 1 e 2”, dentre as duas opções de teste, *percentage split* e *cross-validation*, não se observou uma grande diferença na precisão do modelo treinado. Na primeira opção, o nível geral de acerto foi de

94,7%, ao determinar alunos retidos e perdidos. Na segunda, o teste apresentou um resultado muito próximo, com acerto de 94,8%.

No Quadro 12, observa-se o resultado da matriz de confusão dos dois testes. Nos casos da previsão de perda como "sim", na opção de teste *percentagem split*, o *recall*¹ (0,574) foi ligeiramente menor do que a opção *cross-validation* (0,596).

Quadro 12 – Matriz Confusão ANEAS (2018)

Percentage split (70%)

		PREVISTA PELO MODELO		RECALL
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 54	Falso Negativo 40	0,574
	NÃO	Falso Positivo 22	Verdadeiro Negativo 1054	0,980

Cross-validation (10 folds)

		PREVISTA PELO MODELO		RECALL
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 186	Falso Negativo 126	0,596
	NÃO	Falso Positivo 75	Verdadeiro Negativo 3514	0,979

Fonte: Elaborada pelo autor

Na primeira opção, pode-se interpretar que, da amostra de 94 alunos – selecionados aleatoriamente pelo sistema de teste – que realmente deixaram a instituição, o modelo preditivo obteve 54 casos certos (verdadeiros positivos) e 40 errados (falsos negativos), com um *recall* de 0,574.

Na segunda opção, *cross-validation*, que considera todo o conjunto de dados para teste, dos 312 alunos que deixaram os colégios da ANEAS em 2018, o modelo preditivo indicou a perda de 186 estudantes corretamente (verdadeiro positivo) e 126 erroneamente (falsos negativos), com *recall* 0,596.

Embora a precisão geral do EDS seja aparentemente alta, em torno de 95% de acertos, ele é capaz de interpretar melhor os estudantes que permanecem na instituição, que representam a maior quantidade da amostra, com níveis de quase 97% acertos. Para aqueles que saem, a porcentagem de precisão cai para perto de 60% de acerto.

¹ Porcentagem de dados classificados como positivos comparado com a quantidade real de positivos que existem em nossa amostra. Quanto mais próximo de um maior a assertividade.

De acordo com Han e Kamber (2006), a técnica de *Minimum Spanning Tree* é um método de indução de árvores que permite a remoção de ramos menos confiáveis. Isso permite a indução de árvores menores, menos complexas e mais fáceis de entender. A árvore foi podada para verificar se a precisão poderia ser melhorada. Entretanto, os resultados foram os mesmos ou inferiores quando se utilizaram os 14 atributos. Logo, nos testes realizados, o modelo calculado com todos os atributos se apresentou como o mais eficiente, pois sua taxa de erro foi menor do que os outros modelos com podas treinados.

Para finalizar a análise da qualidade de acertos do modelo com melhor resultado, foram realizadas mais duas verificações. A primeira, em que foi comparada a sua taxa geral de acerto com a taxa de acerto que seria obtida assumindo que todos os estudantes fossem retidos, sem uma análise estruturada (conforme mostra o Quadro 13). Na hipótese apresentada, assumindo a retenção de todos os estudantes, o nível de correção seria de 92,0021% contra 94,8475% do EDS.

Quadro 13 – Comparação matriz confusão EDS com hipótese apresentada

		PREVISTA PELO MODELO		
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 186	Falso Negativo 126	% ACERTOS 59,6154%
	NÃO	Falso Positivo 75	Verdadeiro Negativo 3514	97,9103%
Análise de acertos		ERROS	ACERTOS	% ACERTOS
Cross-validation		201	3700	94,8475%
Hipótese		312	3589	92,0021%
			⬆	2,8454%

Fonte: Elaborado pelo autor

Na segunda verificação, a fim de determinar se a precisão do modelo preditivo muda ao considerar os dados por unidade, novos treinamentos e testes foram realizados com os conjuntos de dados de cada unidade e seus resultados comparados ao modelo com dados ANEAS.

Em outras palavras, para cada conjunto de dados de um colégio, a árvore de decisão foi calculada pelo algoritmo J48, e o modelo foi treinado e testado, desta vez,

utilizando apenas a opção de teste *cross-validation*. O Quadro 14 apresenta os resultados dessa análise.

Quadro 14 – Análise resultados treinamento ANEAS X Colégios

Nome base de dados	RECALL		% Classificações corretas
	Perda sim	Perda Não	
ANEAS	0,596	0,979	94,85%
Anchieta	0,474	0,991	96,22%
Jesuítas	0,846	0,995	98,84%
Loyola	0,725	0,994	96,75%
Sanfra	0,921	0,991	98,32%
Santo Inácio	0,268	0,992	95,24%
São Luis	0,780	0,970	94,56%

Fonte: Elaborado pelo autor

A partir dessa fase, observou-se uma melhora no percentual de classificações corretas dos modelos testados em todas as escolas, exceto uma pequena diminuição no *São Luís*. A precisão da previsão quando o atributo foco PERDA é "sim" cresceu em quatro escolas: *Jesuítas*, *Loyola*, *Sanfra* e *São Luís*. Essa informação significa um maior nível de acerto na identificação de alunos com potencial de saída do colégio. Em contrapartida, o *Santo Inácio* e o *Anchieta* apresentaram um baixo *recall*.

Em conclusão, foi possível observar que a análise individualizada indica que devem ser realizados estudos futuros sobre questões de cada colégio, com o objetivo de aumentar o nível de respostas corretas e, eventualmente, desenvolver EDS específicos para cada colégio.

Retomando o modelo preditivo calculado com os dados de todos os colégios e atributos, ele não apresentou grandes distorções da realidade e obteve um ganho relevante de acerto em relação às análises complementares. Logo, ele foi estabelecido para etapa de validação do modelo preditivo, com o nome de EDS ANEAS.

6.2.2 Treinamento e teste base (2020)

Neste estudo, procuramos investigar os resultados de um modelo preditivo desenvolvido a partir de dados do período pandêmico (2020), analisando sua

qualidade preditiva em comparação com um modelo criado com dados pré-pandêmicos, o EDS ANEAS, referente aos dados de 2018.

Então, antes de avançar para a etapa de validação do modelo proposto, foi realizado o mesmo caminho de treinamento e testes da seção anterior, mas desta vez utilizando o conjunto de dados de 2020, ano em que a pandemia começa no Brasil (Quadro 15).

Quadro 15 – Organização treinamento e teste - Modelo Preditivo período covid-19.

Treinamento e teste	Algoritmo para classificação	Opções de testes	Base dados	Quantidade de instâncias	Observação
3	J48	Percentage split (70%)	2020 (COVID-19)	3999	Divisão randômica pelo Weka de 70% das instâncias para treino e 30% para teste.
4		Cross-validation (Folds 10)			Particionamento pelo WEKA em 10 conjuntos, onde um conjunto será utilizado para treino e outro para teste.

Fonte: Elaborado pelo autor

Os testes do novo modelo preditivo mostraram um nível de acerto de 89,5% e uma taxa de erro de 10,5%; quando 70% das instâncias foram utilizadas para treinamento, e 30% para testes. Na opção de *cross-validation*, o resultado de acerto global foi de 90,8%.

A matriz de confusão, disposta no Quadro 16, mostra que o modelo preditivo construído com dados de 2020 não conseguiu prever nos testes os estudantes com potencial de perda. Na primeira opção de teste, 100% dos estudantes foram classificados como falsos negativos; na segunda opção de teste, quase todos foram classificados como falsos negativos.

Quadro 16 – Matriz Confusão (2020)
Percentage split (70%)

		PREVISTA PELO MODELO		RECALL
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 0	Falso Negativo 126	0,000
	NÃO	Falso Positivo 0	Verdadeiro Negativo 1074	1,000

Cross-validation (10 folds)

		PREVISTA PELO MODELO		RECALL
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 25	Falso Negativo 348	0,067
	NÃO	Falso Positivo 19	Verdadeiro Negativo 3607	0,995

Fonte: Elaborado pelo autor

O desempenho global de acerto é justificado pelo alto *recall*, quando o atributo foco PERDA é “não”, com 1, e 0,996, além de alunos retidos representarem quase a totalidade da amostra. Os resultados dos testes para o modelo preditivo do período pandêmico apresentaram dados imprecisos sobre a perda de estudantes, o que aponta para a necessidade, em projetos futuros, de aprofundar a compreensão das variáveis que levaram a esse estado.

Entretanto, para validação neste projeto, foi retomado o EDS ANEAS, definido anteriormente, que apresentou resultados não distorcidos no contexto geral e com uma margem de acerto aceitável. Na próxima seção, serão demonstrados os resultados desta validação, com a inclusão dos conjuntos de dados de 2019, 2020 e 2021. O objetivo foi simular e validar a predição do EDS ANEAS, classificando dados “futuros” não treinados.

6.3 VALIDAÇÃO EDS

O EDS ANEAS foi validado adicionando sua árvore de decisão, disponível no Apêndice E, ao *Weka* e classificando pelo algoritmo J48, a partir dela os conjuntos de dados não treinados de 2019, 2020 e 2021, com perda "sim" ou "não". Para análise de acertos, foi utilizada a opção de teste *cross validation*.

O Quadro 17 apresenta os resultados da matriz de confusão e *recall*.

Quadro 17 – Matriz confusão (2019 a 2021)

		PREVISTA PELO MODELO		
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 104	Falso Negativo 244	RECALL 0,299
	NÃO	Falso Positivo 12	Verdadeiro Negativo 3510	
2020 Cross-validation (10 folds)				
		PREVISTA PELO MODELO		
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 132	Falso Negativo 241	RECALL 0,354
	NÃO	Falso Positivo 307	Verdadeiro Negativo 3319	
2021 Cross-validation (10 folds)				
		PREVISTA PELO MODELO		
		SIM	NÃO	
DADO REAL	SIM	Verdadeiro Positivo 158	Falso Negativo 229	RECALL 0,408
	NÃO	Falso Positivo 243	Verdadeiro Negativo 3344	

Fonte: Elaborado pelo autor.

De acordo com dados dos anos 2020 e 2021, observou-se uma proporção maior de casos "falsos positivos" do que de casos "verdadeiros positivos". Isso indica que o esforço em eventualmente agir sobre esses estudantes indicados impropriamente como perda potencial deve ser avaliado pela gestão educacional. Ou seja, é preciso ponderar qual a dispersão dos esforços em atuar nesses casos indicados erroneamente, em detrimento do sucesso na retenção daqueles classificados corretamente.

O Quadro 18 apresenta a comparação dos resultados de acerto do EDS ANEAS no teste e nas validações nos anos de 2019 a 2021. O nível geral de acerto foi menor nas validações do que no teste de modelo. O percentual de acerto em 2019 foi de 93,85%, mas teve o menor *recall* (0,299) de perda "sim", em comparação com o resultado do teste (0,596). Isso significa que características específicas do ano de 2019 tornaram o modelo menos preciso. Nos anos de 2018 e 2019, o nível de acerto na validação da perda "sim" foi melhor, mas ainda abaixo do teste.

Quadro 18 – Análise Validação EDS ANEAS, Recall e % Classificação (2019-2021)

EDS ANEAS Cross-validation				
	Nome base de dados	RECALL		% Classificações
		Perda sim	Perda Não	
	Teste	0,596	0,979	94,85%
Validação	2019	0,299	0,997	93,85%
	2020	0,354	0,915	86,27%
	2021	0,408	0,932	88,12%

Fonte: Elaborado pelo autor.

No Quadro 19, é apresentada uma nova análise. Aplicou-se o EDS ANEAS ao conjunto de dados, para validação, dos respectivos colégios nos três anos analisados e observou-se uma variação nas classificações de *recall* e na porcentagem de classificações corretas. Nos colégios e anos destacados em verde, o modelo genérico teve uma taxa de resposta mais assertiva do que nas respostas dos testes da ANEAS.

Os resultados obtidos sugerem que o modelo genérico, o qual considerou os dados de todas os colégios para definir a árvore de decisão, teve uma taxa de acerto maior ao analisar o conjunto de dados específicos do colégio na grande maioria dos casos, sugerindo que, eventualmente, o EDS ANEAS possa atender à maioria dos colégios sem a necessidade de ajustes.

Quadro 19 – Análise validação EDS ANEAS aplicado por colégio (2019 – 2021)

Cross-validation

Nome base de dados	RECALL 2019			RECALL 2020			RECALL 2021		
	Perda sim	Perda Não	% Classificações corretas	Perda sim	Perda Não	% Classificações corretas	Perda sim	Perda Não	% Classificações corretas
ANEAS	0,299	0,997	93,38%	0,354	0,915	86,29%	0,408	0,932	88,12%
Anchieta	0,125	0,994	97,31%	0,172	0,874	81,64%	0,559	0,970	92,77%
Jesuítas	0,273	1,000	96,00%	0,325	0,991	94,49%	0,356	0,715	67,91%
Loyola	0,265	1,000	91,92%	0,429	0,960	90,43%	0,343	0,959	89,60%
Sanfra	0,317	1,000	92,05%	0,500	0,962	91,73%	0,405	0,994	93,40%
Santo Inácio	0,276	0,993	95,27%	0,102	0,923	88,44%	0,255	0,970	93,70%
São Luis	0,355	0,994	88,68%	0,405	0,762	69,94%	0,514	0,972	89,82%

Fonte: Elaborado pelo autor.

Entretanto, para alguns colégios, o *recall* teve um resultado menor, fato que permite inferir que o modelo de previsão genérico não conseguiu capturar as características específicas da unidade. Na aplicação da EDS ANEAS, vale a pena avaliar se o modelo genérico exigirá ajustes quando aplicado ao colégio para um maior ganho de acerto.

Na última análise, a fim de observar a melhoria potencial do poder de previsão do modelo, foi criada uma árvore de decisão, considerando apenas os dados da escola, validada com os respectivos conjuntos de dados. Em outras palavras, foi desenvolvido e validado um EDS para cada colégio, a partir dos seus dados.

O Quadro 20 apresenta a consolidação dos resultados com o *recall* e a porcentagem de classificações corretas, em que as destacadas em verde apresentam melhores resultados do EDS por colégio, comparado com os resultados dos testes do EDS ANEAS.

Tais constatações sugerem que ainda existe a possibilidade de ser desenvolvido um EDS específico do colégio que, em algumas situações, pode eventualmente mostrar melhores respostas do modelo preditivo.

Com base nos resultados das avaliações, pode-se concluir que a proposta do EDS ANEAS, construído a partir de dados administrativos da base ERP Sinergia, foi capaz de atingir um nível de previsão em testes e validação. Constatou-se que a qualidade preditiva do modelo pode ser melhorada ao longo do tempo, organizando os conjuntos de dados, incluindo novos e identificando atributos com maior ganho de informação.

Logo, o modelo tem o potencial de agregar valor à gestão educacional na identificação precoce de perdas, permitindo mais tempo para aplicar estratégias de retenção aos alunos identificados.

Quadro 20 - Análise validação EDS COLÉGIO aplicado por colégio (2019 – 2021)

Cross-validation

Nome base de dados	RECALL 2019		% Classificações corretas	RECALL 2020		% Classificações corretas	RECALL 2021		% Classificações corretas
	Perda sim	Perda Não		Perda sim	Perda Não		Perda sim	Perda Não	
ANEAS	0,299	0,997	93,38%	0,354	0,915	86,29%	0,408	0,932	88,12%
Anchieta	0,125	0,997	97,61%	0,207	0,889	83,33%	0,353	0,943	88,25%
Jesuítas	0,333	0,998	96,17%	0,375	0,965	92,42%	0,525	0,620	61,09%
Loyola	0,255	0,999	91,70%	0,337	0,983	91,60%	0,253	0,984	90,64%
Sanfra	0,341	0,974	90,06%	0,528	0,732	71,20%	0,676	0,969	93,96%
Santo Inácio	0,276	1,000	95,95%	0,163	0,912	87,68%	0,213	0,982	95,67%
São Luis	0,382	0,985	88,37%	0,471	0,688	65,02%	0,622	0,972	91,56%

Fonte: Elaborado pelo autor.

O EDS ainda pode ajudar ao colocar o tema da retenção de estudantes no foco da gestão escolar. Essa visão permitirá que a organização preste mais atenção aos indicadores relevantes, de modo a criar uma cultura de mobilização das estruturas internas para tratar a questão da retenção. Como este estudo apresentou, o tema é de grande relevância para qualquer organização.

6.4 PROPOSTA DE INTERVENÇÃO

Com base na análise dos dados e na estrutura teórica, esta seção apresenta uma proposta para que os colégios da ANEAS identifiquem antecipadamente potenciais perdas de alunos, a fim de otimizar a aplicação de suas estratégias para evitar tal situação.

Neste estudo, ao desenvolver os recursos para análise de dados e construir um modelo preditivo, a base sistêmica para aplicar o modelo à realidade dos colégios foi estabelecida. A implementação foi baseada na alocação de recursos internos da TI Corporativa, da Rede Jesuíta de Educação (RJE) e dos colégios participantes. Isso significa que em princípio não são necessários investimentos adicionais, tais como aquisição de software ou contratação de mão de obra externa.

A Figura 7 apresenta as intervenções propostas, divididas em quatro etapas, a serem implementadas até 2023. Para garantir que esta proposta seja viável, é essencial validar os conceitos com a RJE e estabelecer um engajamento com o planejamento de implantação em suas unidades.

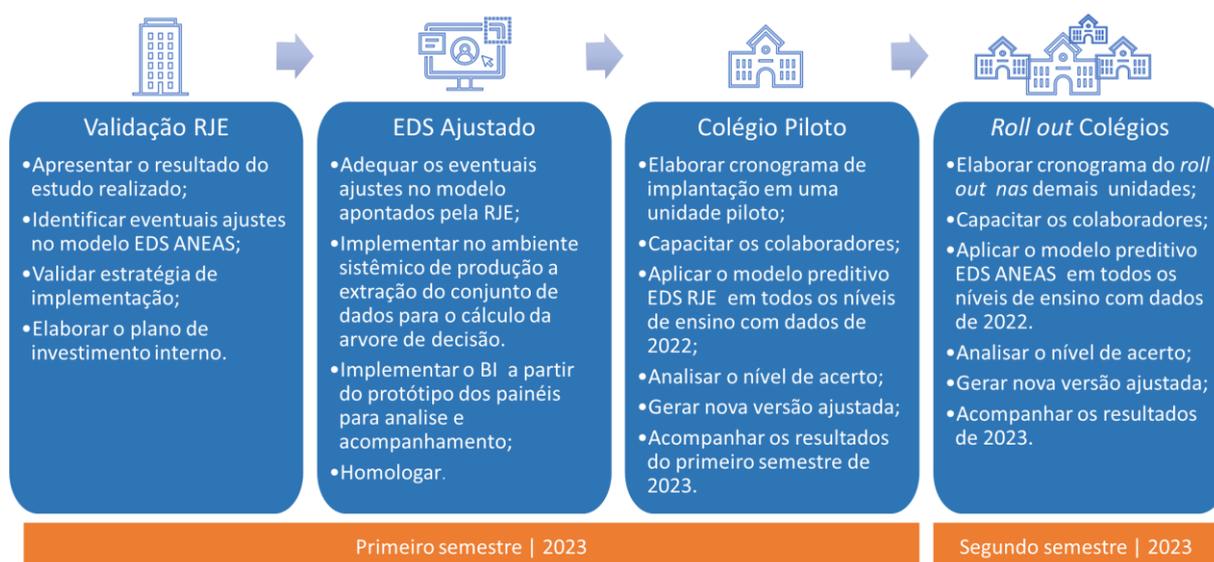
Na primeira etapa, o modelo desenvolvido neste projeto será apresentado à direção da RJE e ao seu conselho consultivo, a fim de detalhar os conceitos e os resultados encontrados. A apresentação incluirá uma explicação detalhada de como

o modelo foi desenvolvido, além de uma lista com todas as variáveis utilizadas e suas definições, assim como uma explicação da funcionalidade do EDS ANEAS na identificação precoce de alunos perdidos.

Uma vez consolidado este entendimento, serão identificados possíveis ajustes que melhorem o EDS ANEAS. Eles serão listados para análise adicional de viabilidade e implementação. O próximo passo será consolidar com a RJE a proposta de implementação do modelo preditivo nos colégios. Entre as verificações, está a estratégia de implementá-lo em uma unidade piloto, definida pela RJE, verificando seu comportamento em operação e as lições aprendidas, antes de implementá-lo nos outros colégios participantes.

Finalizando esta etapa, será necessário preparar o plano de investimento interno, especialmente a alocação de profissionais envolvidos na TI Corporativa e nas unidades participantes, e recolher as aprovações das instâncias responsáveis para oficializar o projeto.

Figura 7 – Proposta de Intervenção



Fonte: Elaborado pelo autor.

Na segunda etapa, o EDS ANEAS será ajustado de acordo com os aspectos apresentados na análise com a RJE, gerando um novo modelo de previsão que será chamado de EDS RJE. Os artefatos desenvolvidos neste projeto de pesquisa (consultas ao banco de dados e painéis de análise no Power BI) deverão ser adequados e implementados no ambiente de produção da ANEAS, buscando a automação da extração dos conjuntos de dados para treinamento, validação e análise.

O desenvolvimento deve ser homologado, assegurando que as extrações e as visões dos instrumentos para análise estejam corretas.

Na terceira etapa, o EDS RJE aprovado pelos assessores da RJE será validado em funcionamento. Para isso, a RJE deverá definir um colégio que possua característica mais próxima dos demais, de modo que um cronograma de implementação piloto possa ser estabelecido. Este organizará atividades que permitirão ao colégio aplicar e analisar os resultados do EDS RJE a seus conjuntos de dados, desta vez considerando todos os níveis de ensino fornecidos, e não apenas os últimos anos do ensino fundamental, como foi feito neste estudo.

A partir dos resultados levantados com dados de 2022, o modelo poderá sofrer eventuais ajustes para a geração da versão final, que será utilizada para acompanhar o conjunto de dados não treinados de 2023. Ao final do primeiro semestre de 2023, o colégio e a RJE analisarão os resultados para encerramento do projeto piloto.

Na quarta e última etapa, em conjunto com o RJE, será elaborado um novo cronograma para a implementação do EDS RJE nos outros colégios participantes. Este cronograma levará em conta as lições aprendidas da etapa anterior. A sequência de atividades será semelhante à do projeto piloto, que consiste em treinamento, aplicação ao conjunto de dados de 2022 e análise dos resultados. Eventualmente, este processo gera uma versão ajustada que será monitorada na operação com dados de 2023.

No primeiro semestre de 2024, a RJE, juntamente com os colégios, analisará se poderá reafirmar se o sistema de detecção precoce (EDS RJE), proposto e aplicado em operação no contexto dos colégios, produz resultado eficiente na detecção precoce da perda de alunos, podendo ser um instrumento de apoio aos gestores educacionais.

7 CONSIDERAÇÕES FINAIS

A perda de alunos é um problema no sistema educacional, devido a várias especificidades locais ou mesmo nacionais. Tal realidade representa desperdícios sociais, acadêmicos e econômicos, pois recursos públicos ou privados são investidos em cada aluno, porém sem o devido retorno. Vale destacar que a problemática também se reflete no desperdício referente a espaço físico, a compra de equipamentos e ao aumento da ociosidade de professores e funcionários. Para o indivíduo, pode levar a consequências negativas a longo prazo, tais como: dificuldades financeiras, baixo nível educacional e maior risco de atividades criminosas.

Klein (2015, p.52) descreve que, no ponto de vista das instituições jesuítas, a perda de um aluno faz com que a escola se distancie do seu objetivo de ajudar no desenvolvimento mais completo possível de todos os talentos dados por Deus a cada indivíduo e de perder a oportunidade da formação humana destes alunos, mais do que a formação acadêmica, que é própria de um colégio jesuíta.

Outro aspecto prático com a perda de alunos nos colégios é o impacto na sustentabilidade financeira. A diminuição da receita das mensalidades pode ser um problema que deve ser analisado com toda atenção. Se os alunos estão saindo, torna-se muito mais difícil para uma escola fazer os movimentos necessários para manter sua missão a longo prazo.

No desenvolvimento deste estudo, o sistema educacional brasileiro foi contextualizado, o que compreende o conjunto de recursos necessários para que o direito à educação seja realizado. Como destaque, referencia-se a Lei nº 9.394, de Diretrizes e Bases da Educação Nacional, de 20 de dezembro de 1996 (LDB 9.394/96), que define e regulamenta o sistema educacional brasileiro, enfatizando que cada instituição pode estabelecer seus próprios padrões de gestão, levando em conta as peculiaridades existentes em cada região e respeitando as organizações e a estrutura dos órgãos responsáveis.

Dessa forma, o fenômeno da perda de alunos representa um desafio para que as escolas o compreendam e tomem medidas para revertê-lo. Manter os alunos garante o cumprimento da sua missão e a sustentabilidade das ações de longo prazo, com mais indivíduos educados com os seus valores e, conseqüentemente, podem trazer impactos positivos para a sociedade.

Posto esse desafio, foi apresentada a conceituação de evasão escolar que significa a saída do aluno do sistema educacional. Nesse contexto, considera-se evasão escolar como perda de alunos. Em outras palavras, ao saírem dos colégios, a gestão escolar não tem como garantir que os alunos que estão sendo acompanhados tenham deixado o sistema ou apenas mudado de instituição.

Também foram compreendidos os principais fatores de perda de alunos, que estão relacionados com características sociodemográficas e socioeconômicas que a escola não pode controlar. Ademais, emergem aspectos práticos da escola que influenciam a aprendizagem e o sucesso escolar, nos quais a escola pode atuar.

Um tema que se apresentou na pesquisa foi o surgimento do novo coronavírus, o qual teve um impacto negativo sobre os sistemas internacionais de saúde e educação. As escolas foram fechadas em muitos países, afetando a capacidade de aprendizagem dos alunos, e muitas pessoas foram isoladas do convívio social. Na esteira da pandemia da covid-19, alguns dos efeitos mais críticos identificados na educação estão associados a processos de ensino-aprendizagem comprometidos e ao aumento das taxas de abandono escolar.

Este estudo, ainda, retoma os conceitos de fidelização de cliente. A partir dos pressupostos do *marketing* tradicional e voltado para instituições educacionais, o foco está na compreensão do ciclo ou do comportamento de compra que influencia as decisões dos pais e alunos de estabelecer um novo contrato de serviço educacional ou manter o vínculo ao longo dos anos. A compreensão deste comportamento permite que a gestão educacional se prepare para agir com os responsáveis acadêmicos, em diferentes estágios do vínculo institucional, com informações que influenciam a decisão de manter seu filho matriculado na instituição.

Dessa forma, a identificação precoce de potenciais perdas estudantis é importante porque permite que estratégias sejam colocadas em prática o mais rápido possível, a fim de reverter o processo de saída do estudante do colégio. O projeto descreve experimentos com sistemas de detecção precoce, chamados EDS, em instituições brasileiras e do exterior, que apresentaram respostas positivas quanto à identificação e à redução da perda de alunos. Foi compreendido que o conceito por trás da construção de um EDS é baseado nos princípios de mineração de dados e da aprendizagem de máquinas. O EDS nada mais é um modelo preditivo definido pelo cálculo de um algoritmo de classificação de dados, que considera atributos relacionados à perda de alunos. Dessa forma, com os dados históricos, é estabelecida

uma árvore de decisão para predição, aplicada a novos conjuntos de dados “futuros” para predição.

Em relação ao primeiro objetivo específico deste estudo, pretendia-se descrever e propor o sistema de detecção precoce de perda de aluno, por meio do EDS. Para sua definição, foram extraídos dados do sistema ERP Sinergia. Isso permitiu uma análise exploratória dos dados, o que ajudou a estabelecer hipóteses porque os alunos abandonaram os últimos anos do ensino fundamental nas escolas jesuítas e quais atributos precisavam ser incluídos no modelo preditivo. Portanto, o objetivo específico foi alcançado.

Na análise exploratória, foi possível ainda identificar descobertas importantes que são registradas no final da seção. Entre elas, estão características que devem ser aprofundadas em estudos futuros, tais como a diminuição da taxa de atrasos de pagamento em 2021, mesmo em um período pandêmico, e uma redução drástica nos registros de ocorrências que se apresentam como atributos importantes para a identificação de perda de alunos.

O modelo proposto, denominado EDS ANEAS, foi desenvolvido a partir dos dados consolidados dos colégios em 2018 e por meio do software *Weka*, que tem como objetivo agregar algoritmos provenientes de diferentes abordagens ou paradigmas na subárea da Inteligência Artificial, dedicada ao estudo de aprendizagem de máquina.

Cabe explicitar, ainda, que a ferramenta de mineração de dados (*Weka*) foi utilizada para realizar as análises computacionais e estatísticas dos dados fornecidos. Esse software tenta, indutivamente a partir dos padrões encontrados, gerar hipóteses de soluções e até mesmo teorias sobre os dados em questão. Os testes do modelo desenvolvido mostraram resultados satisfatórios no nível de precisão, consolidando esta etapa de construção.

Entretanto, observou-se que o modelo preditivo mostrou variações no nível de precisão quando os conjuntos de dados para treinamento (definição da árvore de decisão) e testes foram organizados de forma diferente. Para este fim, o modelo foi testado considerando o conjunto de dados por colégio e, posteriormente, com dados consolidados em 2020 (período pandêmico). Então, deve-se notar que o modelo não é estático, e pode ser ajustado à medida que o pesquisador ganha mais experiência com os fatores de perda do estudante e inclui dados de novos cenários históricos.

O segundo objetivo específico proposto neste estudo foi levantar os resultados dos indicadores estabelecidos no EDS no contexto da ANEAS, avaliando a predição do risco de perda. Ele foi atingido com a utilização da Matriz de Confusão, que estabelece indicadores claros do nível de acerto do EDS ANEAS.

A matriz de confusão é uma forma padrão de avaliar o desempenho de um classificador. Ela apresentou quantas instâncias foram classificadas incorretamente e quantas foram classificadas corretamente nos testes e validações realizadas. Por exemplo, o EDS ANEAS verifica se um determinado estudante de um colégio é identificado ou não como uma perda potencial. Ao usarmos o EDS nos conjuntos de dados de alunos para teste ou validação, ele os classificou como perda “sim” (aluno perdido) e outros como perda “não” (aluno retido).

Foi criada uma matriz de confusão analisando todas as instâncias que são classificadas como aluno perdido e, então, verificou-se qual é a porcentagem dessas que realmente são perda e quais não saíram. O mesmo procedimento foi feito com a classificação de alunos retidos: podemos ver qual a porcentagem daqueles que realmente permaneceram na instituição e quais saíram. Com tais resultados, foi apurado o nível de acerto do EDS ANEAS em diferentes cenários.

O terceiro objetivo específico foi analisar as taxas de acerto na identificação de perfis evasores entende-se que foi atingido por meio das análises feitas nas matrizes de confusão organizadas para os diferentes cenários propostos de validação. A precisão do EDS ANEAS, prevendo 2019, 2020 e 2021 foi menor do que a do conjunto de dados treinados (2018). Isso significa que, apesar dos resultados aceitáveis, há espaço para melhorias. Talvez novos atributos, com maior ganho de informação não identificados neste projeto, possam ser acrescentados em estudos futuros.

Um dos aspectos do EDS ANEAS para todos os anos é que ele apresentou uma resposta mais satisfatória na identificação de alunos retidos do que de alunos perdidos. Tal fato pode ser justificado porque a porcentagem de alunos perdidos é bem menor que a porcentagem de alunos retidos. Dessa forma, o algoritmo se vale de mais casos com características dos alunos retidos para montagem da árvore de decisão.

Na validação do EDS ANEAS, também foi notável que o nível de correção ao aplicá-lo aos dados exclusivos dos colégios gera resultados diferentes. Em alguns casos, o modelo genérico tem uma taxa de acerto maior em alguns colégios do que em outros. Este fato reforça o conceito de que o modelo não deve ser estático e pode

ser adaptado às características de cada colégio. Na aplicação deste modelo nos colégios, pode-se delimitar variáveis locais e de tempo, buscando identificar atributos com maiores ganhos de informação, de acordo com as características locais. Desse modo, permite-se verificar se há uma melhora nas porcentagens de precisão na previsão da perda de alunos.

O último objetivo específico – propor mecanismos sistêmicos e processos para implementação do EDS incluindo como rotina administrativa para apoio aos gestores na elaboração de estratégias para retenção de alunos–, foi cumprido em quase sua integralidade. A partir da revisão da literatura e da construção de recursos para análise dos dados e estabelecimento do modelo preditivo, estabeleceu-se a base sistêmica para aplicação do modelo na realidade dos colégios. O EDS pode ser implantado a partir das soluções sistêmicas corporativas: ERP Sinergia e Power BI. Além disso, foi feito o uso da plataforma *Weka*, sem custos, para construir o modelo preditivo. Entretanto, no projeto não foi possível realizar a automatização da integração dessas soluções.

Neste estudo, foram apresentados protótipos com sugestões de monitoramento e controle que deverão ser revistos na fase de implementação, conforme definido na proposta de intervenção. A avaliação desses protótipos será realizada por uma equipe que inclui representantes tanto do pessoal técnico (TI) quanto do pessoal administrativo (RJE).

A proposta de intervenção definida contempla quatro etapas, que vão da validação do modelo preditivo proposto (EDS ANEAS) com os responsáveis pela RJE, eventuais ajustes no modelo, implementação em uma unidade piloto e lançamento para novas unidades. Ao final desse ciclo, uma vez que os resultados do modelo em operação (EDS RJE) tenham sido consolidados no contexto dos colégios, será possível verificar o nível de precisão e a definição dos procedimentos de acompanhamento e controle.

Logo, observamos neste estudo que o sistema de detecção precoce (EDS), proposto e aplicado no contexto do ensino fundamental, anos finais dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS), produz resultados na detecção precoce da perda de alunos. Em outras palavras, os resultados deste estudo mostram que o EDS proposto oferece informações que subsidiem a detecção precoce da perda de alunos, com a experiência e a conjunção de novos atributos com maior ganho de informação, ele pode ser melhorado.

Como sugestão para futuras pesquisas, recomenda-se concentrar a análise na aplicação de diferentes tipos de algoritmos, avaliando se eles produzem modelos preditivos com resultados mais eficientes. Poderiam, também, ser realizados estudos a fim de determinar quais novos atributos promoveriam um maior ganho de informação, além de verificar a possibilidade de sistematizá-los. Ademais, sugere-se a análise da taxa de evasão escolar nos anos iniciais em comparação com os anos finais do ensino fundamental, a fim de observar mais amplamente os anos com maiores perdas e o comportamento da promoção dos alunos e entre esses grupos.

REFERÊNCIAS

- ALMEIDA, R. E. P.; Kaim, J. Vivências e memórias no espaço escolar, um relato de aprendizagem. **Brazilian Journal of Development**, [S. l.], v. 6, n. 9, p. 65080-65088, 2020. Disponível em: <https://www.brazilianjournals.com/index.php/BRJD/article/view/16129>. Acesso em: 11 jul. 2021.
- ARRUDA, E. P. Educação Remota Emergencial: elementos para políticas públicas na educação brasileira em tempos de covid-19. **Em Rede - Revista de Educação a distância**, Porto Alegre, v. 7, n. 1, p. 257-275, 2020. Disponível em: <https://www.aunirede.org.br/revista/index.php/emrede/article/view/621>. Acesso em: 14 jul. 2021.
- BAYMA, H. F. A. **O abandono escolar no ensino fundamental brasileiro: alguns factores familiares**. [Dissertação] Mestrado. XX f. Faculdade de Psicologia e da Ciências da Educação. Universidade de Coimbra. Coimbra, mai.2008. Disponível em: <https://estudogeral.sib.uc.pt/handle/10316/7512>. Acesso em: 14 abr. 2021.
- BEAN, J. P.; METZNER, B. S. A conceptual model of non traditional undergraduate student attrition. **Review of Educational Research**, Washington, v. 55, n. 4, p. 485-540, 1985.
- BERENS J. *et al.* Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods, **CESifo Working Paper Series**, Munich, n. 7259, 2018. Disponível em: https://ideas.repec.org/p/ces/ceswps/_7259.html. Acesso em: 15 jul. 2021.
- BLACKWELL, Roger D., MINIARD, Paul W., ENGEL, James F. **Comportamento do Consumidor**. 9ed. São Paulo; Cengage. 2008.
- BRASIL. **Lei nº 9.394, de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional. Brasília, DF: Presidência da República, 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 11 jul. 2021.
- BRASIL. Ministério da Educação. Gabinete do Ministro. **Portaria nº 343, de 17 de março de 2020**. Dispõe sobre a substituição das aulas presenciais por aulas em meios digitais enquanto durar a situação de pandemia do Novo Coronavírus -covid-19. Brasília, DF: Diário Oficial da União, mar. 2020a.
- BRASIL. Conselho Nacional de Educação. Conselho Pleno. **Parecer nº 5, de 28 de abril de 2020**. Dispõe sobre a reorganização do Calendário Escolar e da possibilidade de cômputo de atividades não presenciais para fins de cumprimento da carga horária mínima anual, em razão da Pandemia da covid-19. Brasília, DF: Conselho Nacional de Educação, 2020b. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=145011-pcp005-20&category_slug=marco-2020-pdf&Itemid=30192. Acesso em: 15 jul. 2021.

BURGESS, S.; SIEVERTSEN, H. H. Schools, skills, and learning: The impact of covid-19 on education. **VOX CEPR Policy Portal**, 01 abr. 2020. Disponível em: <https://voxeu.org/article/impact-covid-19-education>. Acesso em: 14 jul. 2021.

CERVO, A. L.; BERVIAN, P. A.; DA SILVA, R. Metodologia científica. 6 ed. São Paulo: Pearson Prentice Hall, 2007.

FERREIRA, E. C. S.; OLIVEIRA, N. M. Evasão escolar no ensino médio: causas e consequências. **Scientia Generalis**, [S. l.], v. 1, n. 2, p. 39-48, 2020. Disponível em: <http://scientiageneralis.com.br/index.php/SG/article/view/v1n2a4/15>. Acesso em: 12 jul. 2021.

FERREL, O. C.; HARTLINE, M. D. **Estratégia de Marketing**. São Paulo: Thomson, 2005.

FILHO, J. A. B. L.; SILVEIRA, I. F. **Detecção precoce de estudantes em risco de evasão usando dados administrativos e aprendizagem de máquina**. *Scholarly Journals*, [S. l.], v. 40, p. 480-495, jan. 2021. Disponível em: <https://www.proquest.com/docview/2483973068?pq-origsite=primo>. Acesso em: 15 jul. 2021.

FRITSCH, Rosangela (org). **Ensino Médio: caminhos e descaminhos da evasão escolar**. São Leopoldo: Oikos, 2015.

FUNDAÇÃO LEMANN. **Educação não presencial na perspectiva dos alunos e famílias**. São Paulo: Fundação Lemann, 2020. Disponível em: <https://fundacaolemann.org.br/storage/materials/97r65vVy55x1aZyOUpGhklaAyxYBwtqXTgh5mlb7.pdf>. Acesso em: 14 jul. 2021.

KOTLER, P. **Administração de Marketing** – A edição do novo milênio. 10ª. edição. São Paulo: Pearson Prentice Hall, 2000.

KOTLER, P.; FOX, K. **Marketing Estratégico para Instituições Educacionais** – São Paulo: Atlas, 1994.

KOTLER, P.; KELLER, K. **Marketing Management**. 14th edition. Upper Saddle River, New Jersey: Prentice Hall, 2011.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Métodos e técnicas de Pesquisa Social**. 6 ed. São Paulo: Editora Atlas SA, 2008.

HAN, Jiawei; KAMBER, Micheline. **Data mining Concepts and Techniques**. 2ª Edição. San Francisco: Morgan Kaufmann Publishers, 2006.

HEPPEN, J. B.; THERRIAULT, S. B. Developing early warning systems to identify potential high school dropouts. **National High School Center**, Illinois, p. 1-13, jul. 2008. Disponível em: <https://eric.ed.gov/?id=ED521558>. Acesso em: 15 jul. 2021.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Sinótese Estatística da Educação Básica 2018-21. Inep, 2021. Disponível em < <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados>>. Acesso em: 14 jul 2021.

INSTITUTO SONHO GRANDE. Abandono, evasão escolar e COVID-19. *In: INSTITUTO SONHO GRANDE. Pesquisas em Educação*. [São Paulo, SP]: Instituto Sonho Grande, 2020. Disponível em: http://porvir-prod.s3.amazonaws.com/wp-content/uploads/2021/01/21180400/Sonho-Grande_pesquisa-em-educacao_evasao-e-covid_2020.pdf. Acesso em: 14 jul 2021.

LOPES, N. **Como combater o abandono e a evasão escolar**. Revista Nova Escola. Disponível em: <<http://gestaoescolar.abril.com.br/aprendizagem/como-combater-abandono-evasao-escolar-falta-alunos-abandono-acompanhamento-frequencia-551821.shtml>>. Acesso em: 16 jun 2021.

KEMBER, D. A longitudinal-process model of dropout from distance education. **The Journal of Higher Education**, [S. l.], v. 60, n. 3, p. 278-301, 1989.

MALDONADO-ULLOA, P.Y. *et al.* Sistema de Alerta Temprana, Lineamientos de Operación. **Secretaría de Educación Pública de México**, Programa Síguete, p. 1-18, 2011.

MELO, I. V. **As consequências da pandemia (COVID-19) na rede municipal de ensino: impactos e desafios**. 2020. Trabalho de Conclusão de Curso (Especialização em Docência no Ensino Superior) - Instituto Federal Goiano, Ipameri, 2020. Disponível em: <https://repositorio.ifgoiano.edu.br/handle/prefix/1377>. Acesso em: 15 jul. 2021.

MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. **Machine learning: An artificial intelligence approach**. [S.l.]: Springer Science and Business Media, 2013.

NERI, M. *et al.* **Motivos da evasão escolar**. São Paulo: Instituto Unibanco, 2008. Disponível em: <http://www.bibliotecadigital.abong.org.br/bitstream/handle/11465/1166/1789.pdf>. Acesso em: 11 jul. 2021.

NOROUZI, Monire; SOURI, Alireza; SAMAD ZAMINI, Majid. **A Data Mining Classification Approach for Behavioral Malware Detection**. *Journal of Computer Networks and Communications*, v. 2016, p. 1, 2016. Disponível em: <<http://dx.doi.org/10.1155/2016/8069672>>. Acesso em: 10 abr. 2022.

OREOPOULOS, P.; SALVANES, K. G. Priceless: the nonpecuniary benefits of schooling, **Journal of Economic Perspective**, [S. l.], v. 25, n. 1, p. 159-184, 2011. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.25.1.159>. Acesso em: 15 jul. 2021.

PODESTA, Mora. **El liderazgo de gestión de la calidad en el cambio y la mejora de centros educativos**. [Dissertação] Mestrado. 103 f. Universidad Católica del Uruguay. Montevideo, mai. 2018. Disponível em: <https://www.flacsi.net/wp-content/uploads/2020/01/Tesis-final-3-mayo-word-97.pdf>. Acesso em: 22 nov. 2020.

PORTELA, S. Evasão ou Retenção? Uma questão crucial à sustentabilidade das Instituições de Ensino Superior. *In: JORNADAS REGIONAIS CAMPINAS-SP, 2013*, Campinas, SP. **Anais [...]**. São Paulo: SEMESP, 2013. Disponível em: <http://www.compasso.com.br/clipping/ipc027.pdf>. Acesso em: 12 jul. 2021.

PRETTO, N.L.; BONILLA, M.H.S; SENA, I.P.F.S. (org) Educação em tempos de pandemia: reflexões sobre as implicações do isolamento físico imposto pela COVID-19: Salvador: Edição do autor, 2020.

REDE JESUÍTA DE EDUCAÇÃO. Projeto Educativo Comum. São Paulo: Loyola, 2016.

RIGO, S. J.; CAZELLA, S. C.; CAMBRUZZI, W. Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. **DESAFIE**, [S. I.], p. 168-177, 2012. Disponível em: <http://www.br-ie.org/pub/index.php/desafie/article/view/2787>. Acesso em: 15 jul. 2021.

RIGO, S. J. *et al.* Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, [S. I.], v. 22, n. 1, p. 132-146, 2014. Disponível em: <http://br-ie.org/pub/index.php/rbie/article/view/2423>. Acesso em: 15 jul. 2021.

SAMPIERI, Roberto Hernández; COLLADO, Carlos Fernández; LUCIO, Pilar Baptista. **Metodologia de pesquisa**. 3. ed. São Paulo: McGraw-Hill, 2006.

SANSONE, D. Beyond Early Warning Indicators: High School Dropout and Machine Learning, **SSRN**, [S. I.], 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3062317. Acesso em: 15 jul. 2021.

ŞARA, N-B. *et al.* High-School Dropout Prediction Using Machine Learning: A Danish Large-scale Study. *In: EUROPEAN SYMPOSIUM ON ARTIFICIAL NEURAL NETWORKS, COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING. Proceedings*. Bruges: ESANN, p. 319-324, 2015. Disponível em: http://static-curis.ku.dk/portal/files/168326532/es2015_86.pdf. Acesso em: 15 jul. 2021.

SATHYA, S.; RAJESH, A. **Performance analysis on diabetes prediction with different classification algorithms using Weka**. *International Research Journal in Advanced Engineering and Technology (IRJAET)*, Vol. 1, Issue 4, p 178-190, 2015.

SEIDMAN, A. Spring retention revisited: $RET = EId + (E + I + C)$. **College and University**, [S. I.], v. 71, p. 18-20, 1996.

SENHORAS, E. M. Coronavírus e educação: análise dos impactos assimétricos. **Boletim de Conjuntura**, Boa Vista, v. 2, n. 5, 2020. Disponível em: <https://revista.ufrr.br/boca/article/view/Covid-19Educacao/2945>. Acesso em: 12 jul. 2021.

SCHIFFMAN, Lean G; KANUK, Leslie L. **Comportamento do consumidor**. 9. ed. Rio de Janeiro: LTC, 2009.

SILVA, E. A. **Evasão Escolar**: Uma Análise do Impacto da Proposta Educacional de Apoio ao Desenvolvimento Sustentável (PEADS) em Escolas Públicas em Pernambuco/Brasil. 2018. Dissertação (Mestrado em Ciências da Educação) – Escola Superior de Educação Almeida Garrett, Lisboa, 2018. Disponível em: <https://recil.grupolusofona.pt/handle/10437/8758>. Acesso em: 12 jul. 2021.

SILVA FILHO, Raimundo Barbosa; ARAÚJO, Ronaldo Marcos de Lima. Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. **Educação Por Escrito**, Porto Alegre, v. 8, n. 1, p. 35-48, jan.-jun. 2017.

SILVA, M. J. D. As causas da evasão escolar: estudo de caso de uma escola pública de Ensino Fundamental no município de Acará – PA. **InterEspaço**, Grajaú, v. 2, n. 6, p. 367-378, março. 2017.

SOARES, J.F. E BROOKE, N. (Organizadores). **Pesquisa em Eficácia Escolar, Origem e Trajetórias**. Belo Horizonte: Editora UFMG, 2008.

SOLOMON, Michael R. **O comportamento do consumidor: comprando, possuindo e sendo**. 9. ed. Porto Alegre: Bookman, 2011.

TINTO, V. Dropouts from higher education: a theoretical synthesis of the recent literature. **Review of Educational Research**, Washington, v. 45, p. 89-125, 1975. Disponível em: [https://www.scirp.org/\(S\(vtj3fa45qm1ean45vffcz55\)\)/reference/ReferencesPapers.aspx?ReferencelD=1341353](https://www.scirp.org/(S(vtj3fa45qm1ean45vffcz55))/reference/ReferencesPapers.aspx?ReferencelD=1341353). Acesso em: 15 jul. 2021.

THIOLLENT, Michel. **Metodologia da pesquisa-ação**. São Paulo: Cortez, 1986.

UNESCO. ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **Resumo do Relatório de Monitoramento Global da Educação 2020**: inclusão e educação para todos. Paris: Unesco, 2020.

VAISHNAVI, V.; KUECHLER, W. **Design research in information systems**. 2004. Disponível em: <<http://desrist.org/design-research-in-information-systems>>. Acesso em: 05 mar. 2022.

VITELLI, Ricardo Ferreira. **Políticas públicas estaduais e os indicadores de qualidade do ensino médio: correlações e consequências**. [Tese] Doutorado. 202 f. Universidade Vale do Rio dos Sinos - UNISINOS. São Leopoldo. 2017. Disponível em: http://repositorio.jesuita.org.br/bitstream/handle/UNISINOS/6383/Ricardo%20Ferreira%20Vitelli_.pdf?sequence=1&isAllowed=y. Acesso em: 01 jun. 2020.

WILLIAMS, Graham. **Data mining with Rattle and R**. New York: Springer, 2011.

WITTEN, Ian. H.; FRANK, Eibe. **Data mining: practical machine learning tools and techniques**. 2 ed. San Francisco: Morgan Kaufmann Publishers, 2005.

ZANELLA, L. C. H. Metodologia de pesquisa. 2. ed. reimp. Florianópolis: Departamento de Ciências da Administração/ UFSC, 2013.

APÊNDICE A – GANHO DE INFORMAÇÃO BASE 2018

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_ANEAS**
Instances: 3901
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):

Information Gain Ranking Filter

Ranked attributes:

0.0943614 13 FALTA_MAT
0.0871152 12 FALTA_PORT
0.0397951 11 N_OCORRENCIA
0.0332126 4 RACA_ALUNO
0.0302444 10 OCORRENCIA
0.0060614 9 QDE_PARCELAS
0.0034779 7 DISTORCAO_IDADE
0.0025208 6 IDADE_NOPLETIVO
0.001029 5 POSSUI_DEFICIENCIA
0.0009469 2 TIPO_MATRICULA
0.0006588 1 SERIE
0.0002361 8 ATRASO
0.0000191 3 SEXO_ALUNO

Selected attributes: 13,12,11,4,10,9,7,6,5,2,1,8,3 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_Anchieta**
Instances: 344
Attributes: 14
SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.
Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:
0.07630606 13 FALTA_MAT
0.07006485 10 OCORRENCIA
0.07006485 11 N_OCORRENCIA
0.01815161 1 SERIE
0.0032987 4 RACA_ALUNO
0.00180446 3 SEXO_ALUNO
0.00146129 2 TIPO_MATRICULA
0.00059383 5 POSSUI_DEFICIENCIA
0.00014198 7 DISTORCAO_IDADE
0.00000442 8 ATRASO
0 6 IDADE_NOPLETIVO
0 12 FALTA_PORT
0 9 QDE_PARCELAS

Selected attributes: 13,10,11,1,4,3,2,5,7,8,6,12,9 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_Jesuitas**
Instances: 605
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTRCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.1606422	12	FALTA_PORT
0.1543602	13	FALTA_MAT
0.0184713	11	N_OCORRENCIA
0.0159486	10	OCORRENCIA
0.0073542	1	SERIE
0.0069814	4	RACA_ALUNO
0.0016877	2	TIPO_MATRICULA
0.0009741	8	ATRASO
0.00095	5	POSSUI_DEFICIENCIA
0.000526	7	DISTRCAO_IDADE
0.0000469	3	SEXO_ALUNO
0	9	QDE_PARCELAS
0	6	IDADE_NOPLETIVO

Selected attributes: 12,13,11,10,1,4,2,8,5,7,3,9,6 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_Loyola**
Instances: 922
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.1628105	4	RACA_ALUNO
0.1230465	13	FALTA_MAT
0.1007396	12	FALTA_PORT
0.030132	11	N_OCORRENCIA
0.0267539	10	OCORRENCIA
0.0136032	2	TIPO_MATRICULA
0.0040623	7	DISTORCAO_IDADE
0.0023332	8	ATRASO
0.0004886	5	POSSUI_DEFICIENCIA
0.0001051	1	SERIE
0.0000131	3	SEXO_ALUNO
0	9	QDE_PARCELAS
0	6	IDADE_NOPLETIVO

Selected attributes: 4,13,12,11,10,2,7,8,5,1,3,9,6 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_Sanfra**
Instances: 358
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.37739589	12	FALTA_PORT
0.16896137	11	N_OCORRENCIA
0.12163238	13	FALTA_MAT
0.08694296	10	OCORRENCIA
0.00764653	1	SERIE
0.0044451	2	TIPO_MATRICULA
0.00139666	3	SEXO_ALUNO
0.00090708	7	DISTORCAO_IDADE
0.00090708	5	POSSUI_DEFICIENCIA
0.00050317	4	RACA_ALUNO
0.00000538	8	ATRASO
0	9	QDE_PARCELAS
0	6	IDADE_NOPLETIVO

Selected attributes: 12,11,13,10,1,2,3,7,5,4,8,9,6 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_Santolnacio**
Instances: 1029
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.0409939	12	FALTA_PORT
0.0240595	7	DISTORCAO_IDADE
0.0230922	11	N_OCORRENCIA
0.0207267	6	IDADE_NOPLETIVO
0.0199174	13	FALTA_MAT
0.0170113	10	OCORRENCIA
0.0119872	4	RACA_ALUNO
0.0091561	1	SERIE
0.0039417	5	POSSUI_DEFICIENCIA
0.0024781	2	TIPO_MATRICULA
0.0002243	8	ATRASO
0.0000135	3	SEXO_ALUNO
0	9	QDE_PARCELAS

Selected attributes: 12,7,11,6,13,10,4,1,5,2,8,3,9 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2018_SaoLuis**

Instances: 643

Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):

Information Gain Ranking Filter

Ranked attributes:

0.213739 13 FALTA_MAT
0.183546 11 N_OCORRENCIA
0.161415 12 FALTA_PORT
0.157317 10 OCORRENCIA
0.016897 1 SERIE
0.014274 4 RACA_ALUNO
0.001986 8 ATRASO
0.001537 7 DISTORCAO_IDADE
0.000798 2 TIPO_MATRICULA
0.000786 3 SEXO_ALUNO
0.000306 5 POSSUI_DEFICIENCIA
0 9 QDE_PARCELAS
0 6 IDADE_NOPLETIVO

Selected attributes: 13,11,12,10,1,4,8,7,2,3,5,9,6 : 13

APÊNDICE B – GANHO DE INFORMAÇÃO BASE 2020

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_ANEAS**
Instances: 3999
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:
0.055253 11 N_OCORRENCIA
0.027122 10 OCORRENCIA
0.012136 13 FALTA_MAT
0.010854 4 RACA_ALUNO
0.010374 12 FALTA_PORT
0.005598 2 TIPO_MATRICULA
0.002748 9 QDE_PARCELAS
0.00269 8 ATRASO
0.002639 7 DISTORCAO_IDADE
0.001395 1 SERIE
0.000701 3 SEXO_ALUNO
0.000641 5 POSSUI_DEFICIENCIA
0 6 IDADE_NOPLETIVO

Selected attributes: 11,10,13,4,12,2,9,8,7,1,3,5,6 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_Anchieta**
Instances: 354
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTRCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.030624 7 DISTRCAO_IDADE
0.029826 11 N_OCORRENCIA
0.021265 6 IDADE_NOPLETIVO
0.010038 3 SEXO_ALUNO
0.003549 1 SERIE
0.002954 5 POSSUI_DEFICIENCIA
0.002378 2 TIPO_MATRICULA
0.000942 10 OCORRENCIA
0.000628 8 ATRASO
0.00056 4 RACA_ALUNO
0 12 FALTA_PORT
0 9 QDE_PARCELAS
0 13 FALTA_MAT

Selected attributes: 7,11,6,3,1,5,2,10,8,4,12,9,13 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_Jesuitas**

Instances: 581

Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):

Information Gain Ranking Filter

Ranked attributes:

0.045481 13 FALTA_MAT
0.03633 10 OCORRENCIA
0.03633 11 N_OCORRENCIA
0.010517 1 SERIE
0.010421 6 IDADE_NOPLETIVO
0.006657 5 POSSUI_DEFICIENCIA
0.002981 8 ATRASO
0.002591 4 RACA_ALUNO
0.000303 2 TIPO_MATRICULA
0.000124 3 SEXO_ALUNO
0 9 QDE_PARCELAS
0 12 FALTA_PORT
0 7 DISTORCAO_IDADE

Selected attributes: 13,10,11,1,6,5,8,4,2,3,9,12,7 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_Loyola**

Instances: 941

Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):

Information Gain Ranking Filter

Ranked attributes:

0.082958 12 FALTA_PORT
0.079481 11 N_OCORRENCIA
0.053087 10 OCORRENCIA
0.046819 2 TIPO_MATRICULA
0.030162 13 FALTA_MAT
0.014832 9 QDE_PARCELAS
0.014832 8 ATRASO
0.003152 4 RACA_ALUNO
0.001286 1 SERIE
0.000507 5 POSSUI_DEFICIENCIA
0.000377 3 SEXO_ALUNO
0.000223 7 DISTORCAO_IDADE
0 6 IDADE_NOPLETIVO

Selected attributes: 12,11,10,2,13,9,8,4,1,5,3,7,6 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_Sanfra**
Instances: 375
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTRCAO_IDADE
ATRASSO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:
0.1723033276 11 N_OCORRENCIA
0.141572685 10 OCORRENCIA
0.0178909202 1 SERIE
0.0152658995 7 DISTRCAO_IDADE
0.0076938407 4 RACA_ALUNO
0.0041095322 5 POSSUI_DEFICIENCIA
0.0012174092 8 ATRASSO
0.000176303 2 TIPO_MATRICULA
0.0000000152 3 SEXO_ALUNO
0 9 QDE_PARCELAS
0 6 IDADE_NOPLETIVO
0 12 FALTA_PORT
0 13 FALTA_MAT

Selected attributes: 11,10,1,7,4,5,8,2,3,9,6,12,13 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_Santolnacio**
Instances: 1056
Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):
Information Gain Ranking Filter

Ranked attributes:

0.0208132	11	N_OCORRENCIA
0.0036499	1	SERIE
0.0030344	7	DISTORCAO_IDADE
0.0025611	8	ATRASO
0.0018849	5	POSSUI_DEFICIENCIA
0.001861	4	RACA_ALUNO
0.0005356	10	OCORRENCIA
0.0002853	3	SEXO_ALUNO
0.0000257	2	TIPO_MATRICULA
0	9	QDE_PARCELAS
0	6	IDADE_NOPLETIVO
0	12	FALTA_PORT
0	13	FALTA_MAT

Selected attributes: 11,1,7,8,5,4,10,3,2,9,6,12,13 : 13

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: **EDS_Treino2020_SaoLuis**

Instances: 692

Attributes: 14

SERIE
TIPO_MATRICULA
SEXO_ALUNO
RACA_ALUNO
POSSUI_DEFICIENCIA
IDADE_NOPLETIVO
DISTORCAO_IDADE
ATRASO
QDE_PARCELAS
OCORRENCIA
N_OCORRENCIA
FALTA_PORT
FALTA_MAT
PERDA

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 PERDA):

Information Gain Ranking Filter

Ranked attributes:

0.0811468 4 RACA_ALUNO
0.079032 11 N_OCORRENCIA
0.0105798 10 OCORRENCIA
0.0100501 1 SERIE
0.0025765 7 DISTORCAO_IDADE
0.0019364 3 SEXO_ALUNO
0.0016693 8 ATRASO
0.0000798 2 TIPO_MATRICULA
0 12 FALTA_PORT
0 9 QDE_PARCELAS
0 5 POSSUI_DEFICIENCIA
0 6 IDADE_NOPLETIVO
0 13 FALTA_MAT

Selected attributes: 4,11,10,1,7,3,8,2,12,9,5,6,13 : 13

APÊNDICE C – CLASSIFICAÇÃO EDS

=== Run information ===

Scheme: weka.classifiers.trees.J48 -R -N 3 -Q 1 -M 2

Relation: **EDS_Treino2018_ANEAS**

Instances: 3901

Attributes: 14

SERIE
 TIPO_MATRICULA
 SEXO_ALUNO
 RACA_ALUNO
 POSSUI_DEFICIENCIA
 IDADE_NOPLETIVO
 DISTORCAO_IDADE
 ATRASO
 QDE_PARCELAS
 OCORRENCIA
 N_OCORRENCIA
 FALTA_PORT
 FALTA_MAT
 PERDA

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

Number of Leaves : 35

Size of the tree : 59

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	3700	94.8475 %
Incorrectly Classified Instances	201	5.1525 %
Kappa statistic	0.6216	
Mean absolute error	0.0717	
Root mean squared error	0.1994	
Relative absolute error	48.6869 %	
Root relative squared error	73.5227 %	
Total Number of Instances	3901	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,596	0,021	0,713	0,596	0,649	0,625	0,869	0,663	sim
	0,979	0,404	0,965	0,979	0,972	0,625	0,869	0,973	nao
Weighted Avg.	0,948	0,373	0,945	0,948	0,946	0,625	0,869	0,949	

=== Confusion Matrix ===

a b <-- classified as

186 126 | a = sim

75 3514 | b = nao

APÊNDICE D – VALIDAÇÃO EDS

== Model information ==

Filename: **EDS_ArvoreTreinada_2018_ANEAS.model**

Scheme: weka.classifiers.trees.J48 -R -N 3 -Q 1 -M 2

Relation: EDS_Treino2018_ANEAS

Attributes: 14 [...]

=== Classifier model ===

J48 pruned tree

Number of Leaves : 35

Size of the tree : 59

=== Re-evaluation on test set ===

User supplied test set

Relation: **EDS_Validacao2019_ANEAS**

Instances: unknown (yet). Reading incrementally

Attributes: 14

=== Summary ===

Correctly Classified Instances 3614 93.385 %

Incorrectly Classified Instances 256 6.615 %

Kappa statistic 0.4223

Mean absolute error 0.0852

Root mean squared error 0.2546

Total Number of Instances 3870

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,299	0,003	0,897	0,299	0,448	0,496	0,671	0,379	sim
	0,997	0,701	0,935	0,997	0,965	0,496	0,671	0,939	nao
Weighted Avg.	0,934	0,638	0,932	0,934	0,918	0,496	0,671	0,889	

=== Confusion Matrix ===

a b <-- classified as

104 244 | a = sim

12 3510 | b = nao

=== Re-evaluation on test set ===

User supplied test set

Relation: **EDS_Treino2020_ANEAS**

Instances: unknown (yet). Reading incrementally

Attributes: 14

=== Summary ===

```

Correctly Classified Instances   3451      86.2966 %
Incorrectly Classified Instances  548      13.7034 %
Kappa statistic                 0.2494
Mean absolute error             0.1455
Root mean squared error         0.3226
Total Number of Instances       3999
=== Detailed Accuracy By Class ===

      TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
      0,354  0,085  0,301  0,354  0,325  0,250  0,662  0,198  sim
      0,915  0,646  0,932  0,915  0,924  0,250  0,662  0,936  nao
Weighted Avg.  0,863  0,594  0,873  0,863  0,868  0,250  0,662  0,867
=== Confusion Matrix ===
  a  b <-- classified as
132 241 | a = sim
307 3319 | b = nao

=== Re-evaluation on test set ===

User supplied test set
Relation:  EDS_Validacao2021_ANEAS
Instances:  unknown (yet). Reading incrementally
Attributes:  14

=== Summary ===

Correctly Classified Instances   3502      88.1228 %
Incorrectly Classified Instances  472      11.8772 %
Kappa statistic                 0.3351
Mean absolute error             0.1299
Root mean squared error         0.2995
Total Number of Instances       3974

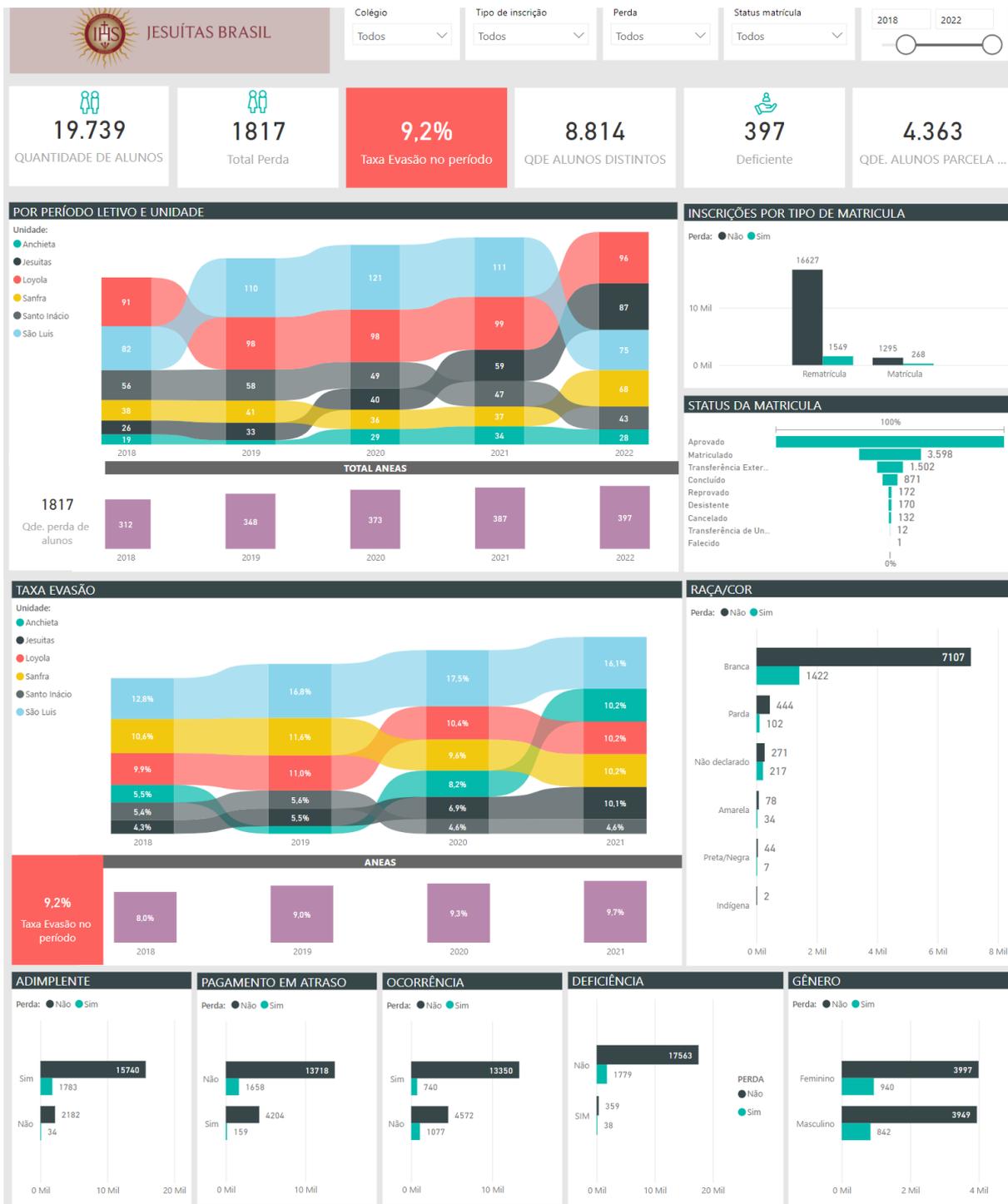
=== Detailed Accuracy By Class ===

      TP Rate FP Rate Precision Recall F-Measure MCC   ROC Area PRC Area Class
      0,408  0,068  0,394  0,408  0,401  0,335  0,724  0,289  sim
      0,932  0,592  0,936  0,932  0,934  0,335  0,724  0,945  nao
Weighted Avg.  0,881  0,541  0,883  0,881  0,882  0,335  0,724  0,881

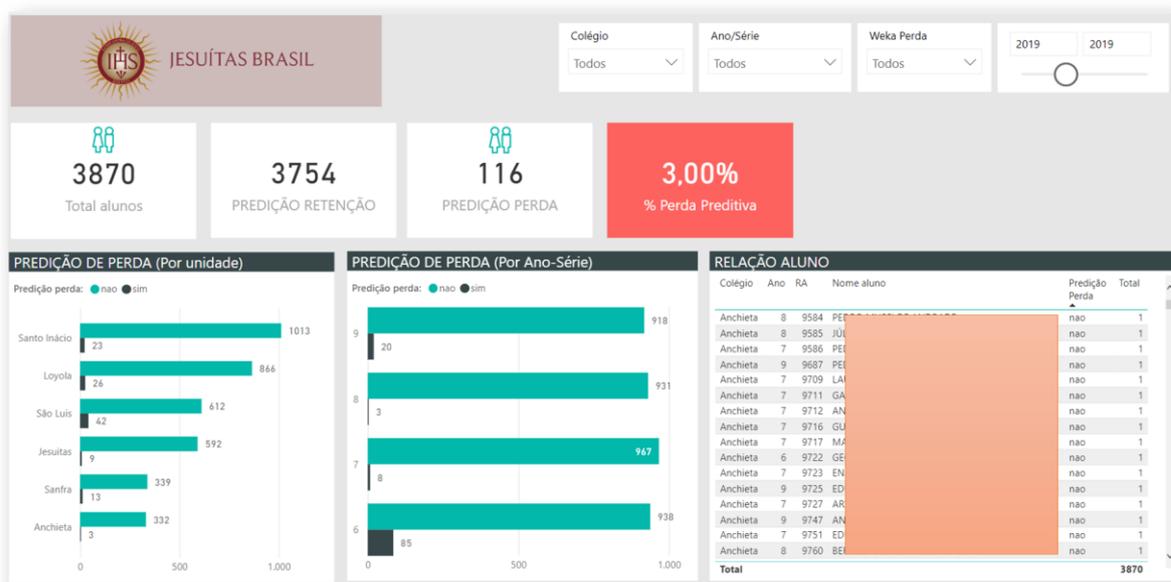
=== Confusion Matrix ===
  a  b <-- classified as
158 229 | a = sim
243 3344 | b = nao

```


APÊNDICE F – BI ANÁLISE PERDA DE ALUNOS



APÊNDICE G – BI PREDIÇÃO WEKA



ANEXO A - CARTA DE ANUÊNCIA

Eu, Fernando Guidini, Diretor Presidente da Rede Jesuíta de Educação, declaro estar ciente de que Breno César Pereira Montalvão efetuará pesquisa intitulada **PERDA DE ALUNOS: Sistema de Detecção Precoce na ANEAS**, no período entre dezembro a julho 2022, com o objetivo de analisar a utilização de dados administrativos e aprendizagem de máquina como instrumentos de apoio à gestão escolar na identificação precoce de perda de alunos do ensino fundamental, anos finais, no contexto dos colégios da Associação Nóbrega de Educação e Assistência Social (ANEAS).

A metodologia prevista consiste na análise dos dados institucionais do ensino fundamental, anos finais, de seis colégios da mantenedora ANEAS, a partir das informações acadêmicas e financeiras contidas no Sistema de Gestão Integrado Sinergia. Desta forma, o mestrando terá acesso aos relatórios e desenvolvimento de consultas ao banco de dados extraindo elementos relacionados ao objeto da pesquisa.

Estando esta instituição em condições para o desenvolvimento deste projeto, autorizo sua execução.

São Paulo, 01 de janeiro de 2022.



Fernando Guidini

Diretor Presidente da Rede Jesuíta de Educação