Monitum - Um Modelo de Classificação para a Predição da Evasão Discente no Ensino Médio

Jônatas Silva Santos
Instituto Federal da Bahia - Campus Salvador
Bahia - Brasil
Email: jonattasts@gmail.com

Pablo Florentino
Instituto Federal da Bahia - Campus Salvador
Bahia - Brasil
Email: pablovf@ifba.edu.br

***Resumo*—A evasão escolar é, certamente, um dos problemas que mais afetam as instituições de ensino em geral. A busca de suas causas tem sido objeto de muitos trabalhos e pesquisas educacionais. As perdas de estudantes que iniciam, mas não terminam seus cursos, geram danos sociais, acadêmicos e econômicos. No setor público, busca-se aprimorar a assistência estudantil e as práticas pedagógicas no sentido de acolher os alunos que possuam realidades sociais adversas e aumentar o tempo de permanência destes nas instituições públicas de ensino. No setor privado, é uma perda relevante de receitas. Em ambos os casos, a evasão acaba impactando não só os alunos como também os professores, funcionários e o uso de equipamentos e do espaço físico. Neste contexto, este trabalho visa aprofundar a elaboração do modelo de classificação que foi utilizado como instrumento para a identificação de potenciais casos de alunos em risco de evasão, com o propósito de orientar o trabalho de profissionais da área pedagógica.**

***Keywords*—Evasão Escolar, Modelo de Predição, Mineração de Dados, KDD - Knowledge Discovery in Database, Análise de Dados.**

I. Introdução

O processo de tomada de decisão é sempre complexo, pois influencia diretamente no futuro das organizações. As instituições de ensino estão dentro deste contexto. Sob o enfoque da gestão do ensino, dois aspectos merecem especial atenção dos dirigentes: Evasão Escolar e Tempo de Permanência dos alunos. Há uma escassez nas Instituições de Ensino de um instrumento avaliador e analítico que possa contribuir com o acompanhamento dos sintomas da evasão discente. É através da disponibilização de uma ferramenta que especialistas da área pedagógica obterão apoio na tomada de decisão e no suporte ao planejamento estratégico. Com o uso desse tipo de ferramenta, novas medidas poderão ser tomadas frente às evidências encontradas.

A evasão escolar acontece quando o aluno abandona a escola, deixando assim, de frequentar as aulas, fazendo com que o seu conhecimento não seja construído adequadamente. Ela não é um problema somente do aluno que deixa a escola, mas sim de toda a sociedade. A evasão escolar pode ocorrer por motivos internos (acadêmicos) ou externos (sociais e econômicos) à escola. Dessa forma, é necessário conhecer a gênese do problema para poder auxiliar a instituição na tomada decisão. Para combater ou controlar a evasão escolar em uma escola é necessário identificar os problemas que afetam os alunos para, assim, buscar novos caminhos, novas estratégias e desenvolver projetos e técnicas pedagógicas que os estimulem a permanecer [1].

O problema do grande número de alunos evadidos em cursos do ensino médio pode ser amenizado ao utilizar sistemas analíticos que deem suporte a especialistas em identificar potenciais casos de alunos propensos à evasão. Para isso, a análise dos padrões de mineração deve estar baseada em informações de profissionais da área pedagógica ao invés de utilizar características genéricas para que não haja uma padronização de um perfil evasor segregado. É possível perceber essa evasão também no ensino superior afetando até Instituições Federais, como mostram os autores [2] [3]. Na prática, o uso de sistemas analíticos e a utilização das técnicas de descoberta de conhecimento (*KDD*) em ambientes educacionais tornam-se ferramentas importantíssimas para municiar a gestão de uma instituição de ensino nos quesitos de gerenciamento e tomada rápida de decisões para resgatar os alunos com dificuldades e intervir na vida acadêmica ao notar um baixo rendimento através de padrões associados a aspectos relacionados ao aluno [4].

Neste contexto, o processo de tomada de decisão tornou-se complexo pela necessidade de informações externas e internas. As informações externas, normalmente, não são estruturadas e são difíceis de serem capturadas. As informações internas tendem a ter uma maior facilidade de serem obtidas quando comparadas às primeiras, dependendo do seu estágio de automação. Mesmo para empresas que possuam sistemas de informação automatizados, a forma de concepção e o objetivo destes, frequentemente, não atendem à área estratégica, pois foram sistemas concebidos para atender a área operacional que trabalha com dados, divergindo da necessidade da área estratégica que trabalha com conhecimento. O processo de transformação de dados em conhecimento faz com que os profissionais de educação possam tomar decisões mais embasadas.

Considerando esta lacuna, o presente trabalho foi desenvolvido a partir de uma pesquisa sobre evasão escolar realizada no IFBA/Campus Porto Seguro, da qual participamos como equipe de desenvolvimento de um produto educacional (Monitum) na área da Educação Profissional e Tecnológica [5]. Essa ferramenta foi utilizada para quantificar os dados gerados pela pesquisa, servindo como um instrumento auxiliar na tomada de decisões para as intervenções realizadas pela equipe pedagógica multidisciplinar do Campus. Apesar dessa proposta inicial, seu uso pode ser ampliado, pois o combate a evasão deve envolver toda a comunidade acadêmica que, em comum acordo, partilha do mesmo objetivo. A partir do trabalho

anterior, busca-se, no presente estudo, aprofundar a elaboração desse modelo de classificação, que visa a identificação precoce do problema, antecipando as ações com o propósito de evitar a evasão escolar.

Para o processo de concepção deste trabalho, foi necessária uma revisão bibliográfica tendo como objetivo o levantamento da fundamentação teórica do tema proposto. O objetivo principal deste trabalho é o desenvolvimento de uma ferramenta que permita identificar casos potenciais de evasão, de acordo com as características definidas por profissionais da área pedagógica. Como objetivos específicos: Identificar padrões através da descoberta de conhecimento sobre os dados obtidos a partir da experiência e do conhecimento de profissionais da área pedagógica, auxiliar na tomada de medidas socioeducacionais que ampliem a permanência dos alunos dos cursos integrados ao ensino médio do IFBA e evitar a criação de um perfil evasor definido exclusivamente por um aspecto social ou econômico.

Inicialmente, para testar o sistema na identificação dos alunos com maior risco de abandono escolar, foram lançados os dados dos estudantes matriculados nos primeiros anos dos Cursos Técnicos Integrados do Campus Porto Seguro. Os cursos técnicos integrados são cursos voltados para alunos que concluíram o ensino fundamental e proporcionam a formação básica, correspondente ao ensino médio, integrada à formação profissional. Segundo estudos, é na primeira série desses cursos que se concentram os maiores percentuais de reprovação e evasão escolar. Na pesquisa realizada para a implantação desse sistema, entre os anos de 2015 e 2018, mais de 60% dos alunos que evadiram no Campus Porto Seguro o fizeram enquanto cursavam o 1º ano [5]. Depois dessa primeira experiência, de acordo com os resultados obtidos, o modelo seria aperfeiçoado diante das demandas que porventura surgissem no período de testagem e, então, seria implantado gradualmente nas outras séries.

Portanto, na próxima seção será apresentado os principais assuntos relacionados a este trabalho. A subseção II-A apresenta o conceito de Evasão Escolar e seu impacto no âmbito acadêmico e social; A subseção II-B aborda a noção de Mineração de Dados. A subseção II-C aponta a conceituação respectiva ao (*KDD - Knowledge Discovery in Database*). A subseção II-D apresenta conceitos relacionados a Análise de Dados. A subseção II-E traz conceitos relacionados a Sistema de Predição.

II. Fundamentação Teórica

Para entender o contexto no qual o trabalho está situado, é importante conhecer alguns conceitos em que o mesmo se fundamenta. Por isso, esta sessão tem por objetivo pautar os trabalhos e tecnologias que serviram de referencial para o mesmo.

A. Evasão Escolar

Segundo Lobo [3], esse fenômeno pode ser medido em uma instituição de ensino superior ou ensino médio, em um curso, em um período de oferta de cursos e em qualquer outro universo de ensino, desde que tenhamos acesso a dados e informações pertinentes. Já o estudo interno, realizado por uma Intituição de Ensino com base em seus dados, pode ser muitas vezes mais detalhado porque é possível institucionalizar-se um mecanismo de acompanhamento da evasão, registrando os diversos casos, agrupando e analisando subgrupos ou identificando situações distintas da evasão, tais como cancelamento, trancamento, transferência ou desistência.

De acordo com Dore [6], alguns pesquisadores distinguem três dimensões conceituais indispensáveis à investigação do abandono escolar:

1) Níveis de escolaridade em que ela ocorre: ensino fundamental, médio ou superior;
2) Tipos de evasão: descontinuidade, interrupção temporária, a não-conclusão definitiva, dentre outras;
3) Razões que motivam a evasão: a escolha de outra escola, necessidade de trabalho, doenças, problemas na escola, problemas pessoais ou problemas sociais.

Salienta-se que, a despeito de qualquer uma das três dimensões, a evasão escolar gera consequências e prejuízos tanto para o estudante, no âmbito pessoal, quanto para sua família e sociedade. Contudo, ao se considerar o momento da trajetória escolar em que ocorre esse abandono, suas implicações são muito mais danosas caso ocorra antes mesmo da conclusão da educação básica. No Brasil, a etapa em que há maiores taxas de reprovação e evasão escolar é o ensino médio. Em 2018, cerca de 7% do total de matriculados nesse nível interromperam seus estudos, o que representa mais de 460 mil estudantes. Segundo esse mesmo estudo, realizado pelo IBGE, a interrupção precoce dos estudos pode elevar as possibilidades do jovem de 18 anos não estar frequentando uma instituição de ensino tampouco inserido no mercado de trabalho [5].

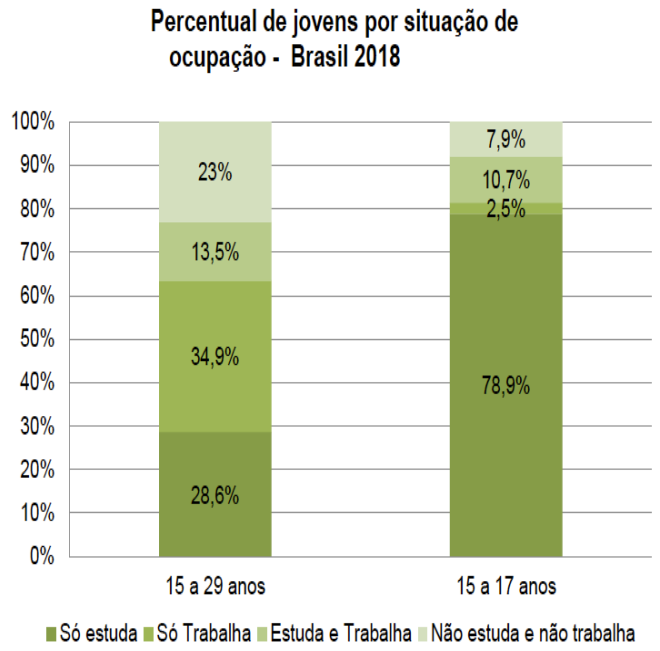

Figura 1: Percentual de jovens por situação de ocupação - Brasil 2018. Fonte: [5]

A Síntese de Indicadores Sociais (*SIS*) do IBGE destaca que, quanto menor o nível de instrução dos jovens que não estudam nem trabalham, maior a prevalência de indivíduos fora

da força de trabalho. Esses jovens não procuravam trabalho no período de referência ou não estavam disponíveis para começar a trabalhar se surgisse uma vaga. Entre os jovens adolescentes de 15 a 17 anos, que corresponde à população potencial do ensino médio, o percentual dos que não estavam frequentando uma instituição de ensino, nem inseridos no mercado de trabalho foi de 7,9%, conforme demonstrado na Figura 1 [5].

No estudo "Enfrentamento da cultura do fracasso escolar"[7], publicado em janeiro de 2021, o Fundo das Nações Unidas pela Infância (Unicef) estima que aproximadamente 4,1 milhões de crianças e adolescentes de 6 a 17 anos tiveram dificuldade de acesso ao ensino remoto em 2020. E que cerca de 1,3 milhão abandonou a escola. Os dados usados no relatório são da Pesquisa Nacional por Amostra de Domicílios (Pnad) de outubro de 2020 [8].

São variadas as motivações para os estudantes abandonarem sua trajetória escolar e é fundamental compreendê-las para que sejam implementadas ações mais efetivas para o enfrentamento do problema. Assim, a evasão escolar é o resultado da junção de várias dimensões – sociais, culturais, políticas, econômicas, pedagógicas – que coexistem nas instituições de ensino.

Outro aspecto considerado significativo no estudo da evasão escolar está correlacionado à perspectiva adotada no exame do problema: sob a ótica do indivíduo, da escola ou do sistema de ensino. Algumas vezes, o que é entendido pelo sistema como um problema de evasão, não o é para o indivíduo ou para a escola, quando considerado isoladamente [6]. Sejam singulares ou interligados, esses aspectos formam entraves que impedem que os indivíduos tenham acesso a uma formação qualificada e que os possibilitem atuar no desenvolvimento da sociedade.

Assim, os comportamentos e atitudes manifestados pelos estudantes já podem estar relacionados à evasão escolar, visto que demonstram um maior ou menor engajamento com a vida escolar. Apesar de existirem várias teorias sobre a temática, em sua maioria, elas reiteram a existência de dois tipos principais de engajamento escolar: o engajamento acadêmico ou de aprendizagem, e o engajamento social ou de convivência do estudante com os colegas, com os professores e com os demais indivíduos da comunidade escolar. A maneira que o estudante lida com esses dois aspectos, ou seja, quanto mais envolvido no ambiente escolar, reduzem-se suas possibilidades de evasão [9].

Em [10] são ressaltadas cinco categorias de abandono:

1) *Pushout* — estudantes "indesejáveis"que a escola procura ativamente afastar;
2) *Disaffiliated* — estudantes que não demonstram vínculos com a escola;
3) *Educational mortalities* — estudantes que não conseguem completar o ciclo de estudos;
4) *Capable dropout* — estudantes cujas capacidades são adequadas aos programas escolares, mas não conseguem se adaptar as demandas da escola;
5) *Stop-Out* — estudantes que deixam a escola e ficam fora por um breve período para depois voltarem.

Nesse sentido, quando o aluno evade, de fato, da escola, ele provavelmente já passou por um longo processo evolutivo quanto ao abandono, mesmo que ainda estivesse assistindo às aulas. O ato de deixar de frequentar definitivamente a escola é somente a aceitação de uma decisão que começou a ser pensada bem antes. É bem provável que esse estudante já tivesse demonstrado com ações e atitudes quanto à possibilidade de desistência.

Seguindo esse parâmetro, é possível identificar precocemente a possibilidade de ocorrência da evasão, visando o acompanhamento individual daqueles em situação de risco. É nesse sentido que a ferramenta desenvolvida se propõe a atuar: fornecer dados que ajudem os profissionais a se antecipar à evasão escolar, traçando ações voltadas a esses estudantes mais propensos a interromper seus estudos.

B. KDD - Knowledge Discovery in Database

KDD refere-se ao processo geral de descoberta de conhecimento útil a partir de dados. A tomada de decisão sobre o que se qualifica como conhecimento envolve a avaliação e, possivelmente, a interpretação dos padrões. Antes da etapa de mineração de dados também se inclui a escolha de esquemas de codificação, pré-processamento, amostragem e projeções dos dados [11]

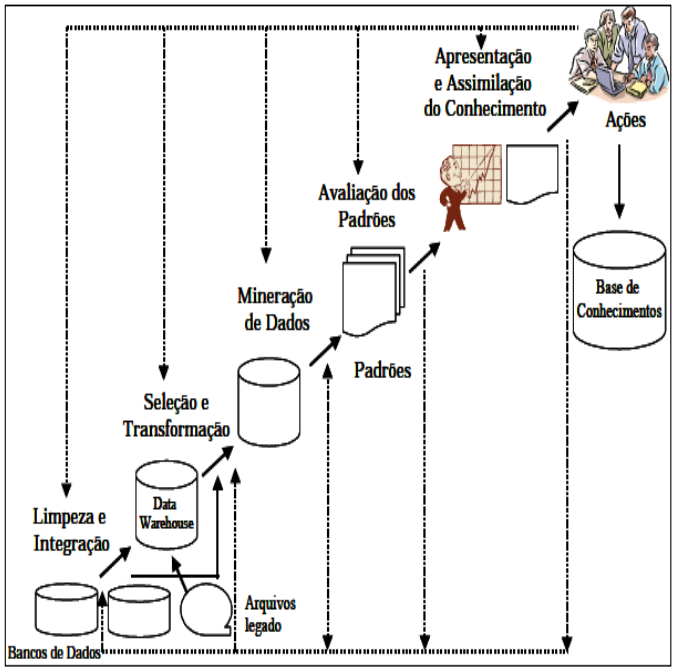

Figura 2: Etapas da descoberta do conhecimento. Fonte: [12]

Portanto como demonstrado na figura 2 e também é apresentado em Costa [12], o processo de KDD consiste de uma sequência iterativa de passos:

1) Limpeza de dados - remove dados inconsistentes e fora dos padrões (noise data);
2) Integração de dados - possibilita a integração de várias fontes de dados, mantendo a consistência e coerência dos dados integrados;
3) Seleção dos dados - seleciona os dados relevantes para aplicação das técnicas de mineração de dados;

4) Transformação de dados – possibilita a transformação ou consolidação dos dados no formato apropriado para o processo de mineração (mining), através de operações do tipo sumarização ou agregação, entre outras técnicas;
5) Mineração dos dados – processo essencial, onde técnicas são aplicadas para análise e extração de padrões dos dados;
6) Avaliação dos Padrões – identifica os padrões verdadeiramente interessantes entre os diversos apresentados pelo processo de mineração de dados, baseados em algumas medidas de interesse;
7) Apresentação e assimilação do conhecimento – utiliza técnicas de visualização e representação do conhecimento para apresentar o conhecimento adquirido aos usuários, bem como introduzi-los no âmbito estudado.

C. Mineração de Dados

Como o termo indica, mineração de dados refere-se á mineração ou descoberta de novas informações em termos de padrões, ou regras com base em grandes quantidades de dados. O objetivo de um armazém de dados (*data warehouse*) é dar suporte à tomada de decisão com dados. Em certos tipos de decisões a mineração de dados pode ser usada junto com um data warehouse. Ela também pode ser aplicada a bancos de dados operacionais com transações individuais, pois ajuda na extração de novos padrões significativos que não podem ser necessariamente encontrados apenas ao consultar ou processar dados, ou metadados no data warehouse [13].

Nas últimas décadas, a mineração de dados cresceu exponencialmente, permitindo a extração de informações relevantes em grandes quantidades de dados. A mineração de dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, buscam identificar um conjunto de padrões de certos dados [4].

A mineração de dados pode prever comportamentos futuros, usando técnicas para identificar a existência de um item, um evento ou uma atividade, ou até mesmo particionando os dados de modo que diferentes classes ou categorias possam ser identificadas com base em combinações de parâmetros [13].

Uma metodologia pode estar baseada em mineração de dados como em [14] onde foi usada para o acompanhamento e identificação precoce dos estudantes com grande potencial de desistência ou desligamento compulsório. Essa abordagem utilizou exclusivamente o histórico de desempenho nas disciplinas do primeiro ano do curso, dispensando fontes externas de dados mais escassas ou de difícil obtenção. Tais instrumentos puderam balizar a tomada de ações individuais direcionadas a alunos em risco, bem como o planejamento de ações futuras.

D. Análise de Dados

Em 1962, Turkey, pensou a Análise dos Dados como, dentre outras coisas, um procedimento ou forma de planejar a coleta de dados para tornar sua análise mais exata e precisa, onde todo um maquinário e resultados de estatísticas deveriam ser aplicadas à análise dos dados [15].

Não muito diferente dos dias atuais, estudiosos acreditam que um projeto de análise de dados consiste em um ciclo composto por seis fases — 1) Descoberta, 2) Preparação, 3) Planejamento do Modelo, 4) Construção do Modelo, 5) Comunicação dos Resultados e 6) Operacionalização [16].

1) **Descoberta** — A primeira fase consiste no momento em que hipóteses iniciais costumam ser levantadas pelo pesquisador, podendo as mesmas serem testadas, posteriormente, com os dados atualmente disponíveis. Deve-se aprender o máximo possível sobre as fontes de dados necessárias e disponíveis para o projeto em que se quer construir; avaliar os recursos disponíveis, que podem incluir tecnologia, ferramentas, sistemas e pessoas; identificar as principais partes interessadas no projeto, que devem incluir qualquer pessoa que se beneficiará com o projeto ou será significativamente impactada por ele. Portanto, é de suma importância que nesta fase o pesquisador adquira o máximo de conhecimento sobre seus dados e o problema que deseja resolver;
2) **Preparação** — A fase de preparação dos dados costuma ser a fase onde pesquisadores gastam, ao menos, 50% do seu tempo. Aqui é necessário explorar, limpar, padronizar, organizar e condicionar os dados antes de preparar sua modelagem para futuras análises. Nesta etapa, costuma-se remover colunas irrelevantes ao projeto, identificar outliers (valores atípicos, que fogem da normalidade), quantidade de colunas com variáveis nulas, realizar relacionamentos entre outras tabelas, etc. Em alguns casos, a visualização dos dados também pode ser realizada com o objetivo de ajudar na compreensão dos mesmos;
3) **Planejamento** — A terceira fase se refere ao planejamento de um modelo onde equipes determinam os métodos, técnicas e fluxos do trabalho que se pretende seguir durante a fase de construção de um modelo. Aqui, também, a equipe explora os dados para aprender sobre os relacionamentos que existem entre as variáveis e, consequentemente, seleciona variáveis-chave e os modelos mais adequados;
4) **Construção** — Na construção, desenvolvem-se os bancos de dados para teste, treino e ambiente de produção. Ademais, nesta fase as equipes constroem e executam modelos baseados no trabalho realizado na fase anterior. A equipe também considera se existem ferramentas que serão suficientes ou adequadas para executar os modelos, ou se será necessário um ambiente mais robusto para execução dos mesmos;
5) **Comunicação** — Nesta etapa os resultados são avaliados a partir dos critérios desenvolvidos na primeira fase, em colaboração com as partes interessadas do projeto. As principais descobertas devem ser identificadas, o valor do negócio quantificado e uma narrativa deve ser desenvolvida para resumir e comunicar as descobertas;
6) **Operacionalização** — Na operacionalização são entregues relatórios finais, orientações, código e documentação técnica. Também poderá ser executado um projeto-piloto.

Ainda em um projeto de Análise de Dados, encontramos quatro tipos de estruturas de dados — Estruturados, Semiestruturados, Quase estruturados e Não Estruturados [16].

- **Estruturados** — Dados com estrutura muito bem definida como arquivos CSV, Planilhas e Sistemas de Gerenciamento de banco de dados relacional.
- **Semiestruturados** — Dados textuais que costumam seguir algum padrão, como arquivos XML.
- **Quase Estruturados** — Dados textuais que normalmente necessitam de esforço, ferramentas e tempo para serem trabalhados.
- **Não Estruturados** — Dados sem estruturas, como documentos de texto, PDF's, imagens e vídeo.

E. Sistemas de Predição

Um sistema preditivo é baseado técnicas de descoberta de conhecimento capazes de identificar padrões por meio de uma série de dados previamente coletados. Nem sempre esses dados são fáceis de serem encontrados e, para isso, é comum fazer uso de algoritmos para obtê-las em razão da grande capacidade computacional existente. Um modelo preditivo serve para reconhecer e identificar padrões dentro de um conjunto de dados auxiliando na realização da análise de dados [17].

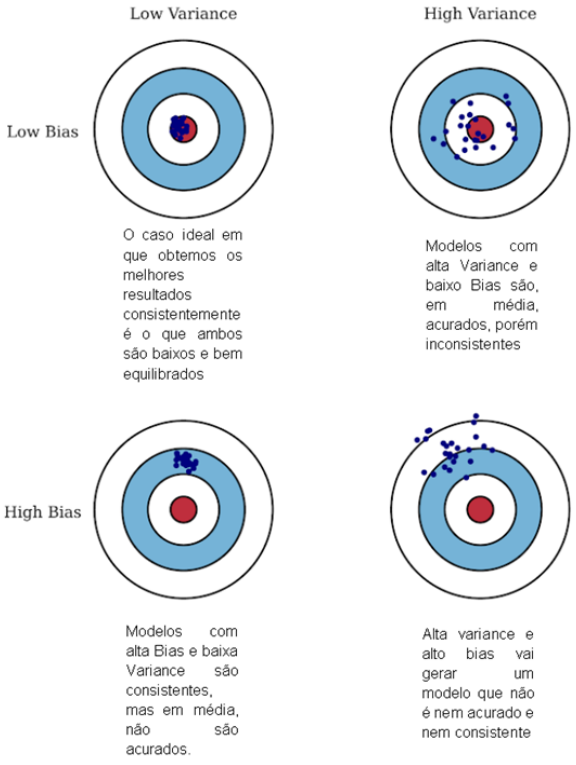

Figura 3: Cada ponto azul representa um modelo e o centro vermelho representa o ponto ideal. Fonte: [17]

Um fator que pode causar erro na predição é chamado de viés (*bias*), que é a diferença entre a predição média do modelo preditivo e o valor correto esperado. Sendo assim, um sistema com viés aprende relações erradas e gera previsões inconsistentes. O sistema não aprende corretamente com o conjunto de dados, assumindo muitas informações sobre os dados que não são necessariamente corretas como foi possível observar na figura 3 [18].

Uma solução preditiva está propensa a alcançar êxito no âmbito de definir padrões e encontrar possíveis casos de evasões estudantil. Porém o viés discriminatório é um fator que deve ser considerado e que pode inviabilizar o uso confiável de um sistema de predição no combate a evasão. Um exemplo disso é descrito em [19] onde o modelo matemático alternativo aos exames de seletividade, adotado em função da pandemia pelo Governo Britânico foi considerado injusto ao utilizar exclusivamente o histórico de desempenho nas disciplinas do primeiro ano do curso.

III. Trabalhos Correlatos

Abaixo são apontados alguns trabalhos que correlacionam-se, em maior ou menor grau, com algumas das funcionalidades deste projeto, ou serviram de inspiração na concepção e no andamento deste trabalho.

Em [20] é feita uma análise multidimensional do perfil socioeconômico dos alunos com o objetivo de auxiliar na análise da possível influência deste em relação ao desempenho acadêmico e a evasão escolar. Para análise do perfil do aluno o trabalho utiliza técnicas de datawarehouse e mineração de dados sobre as informações coletadas através de uma pesquisa espontânea realizada no Instituto Federal da Bahia - Campus de Salvador. Além disso a pesquisa está voltada a indicar novos fatores a serem avaliados como mostra a figura 4 com o intuito de auxiliar na construção de ações afirmativas de assistência estudantil visando aumentar a permanência dos alunos nos cursos.

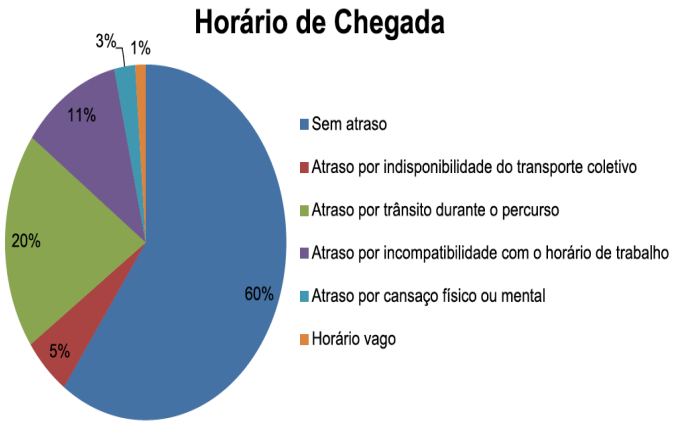

Figura 4: Distribuição dos motivos de chegada com atraso pelos alunos. Fonte: [20]

O SDBayes [21] é um software que tem o intuito de auxiliar a gestão universitária especialmente com o problema de evasão discente. A técnica utilizada para analisar os dados foram as Redes Bayesianas que através do método de aprendizagem *Expectation Maximization* (EM) onde foram analisados os alunos dos cursos de Ciência da Computação, Administração e Engenharia de Produção do ensino superior. Foram consideradas 9 variáveis como parâmetros como mostra a figura 5 para calcular a probabilidade do aluno evadir. Sobre o fator do viés não foi encontrada abordagem no trabalho sobre como a ferramenta lida com essa possibilidade.

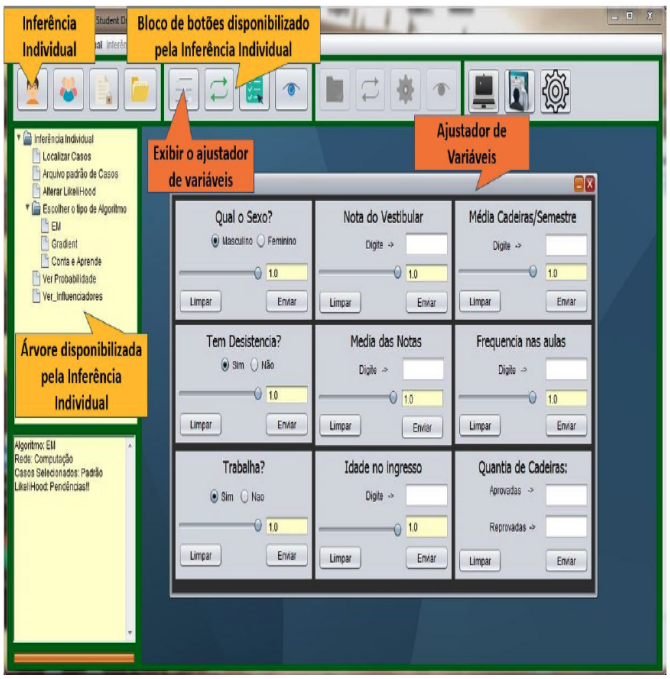

Figura 5: Preenchimento das variáveis na ferramenta SDBayes. Fonte: [21]

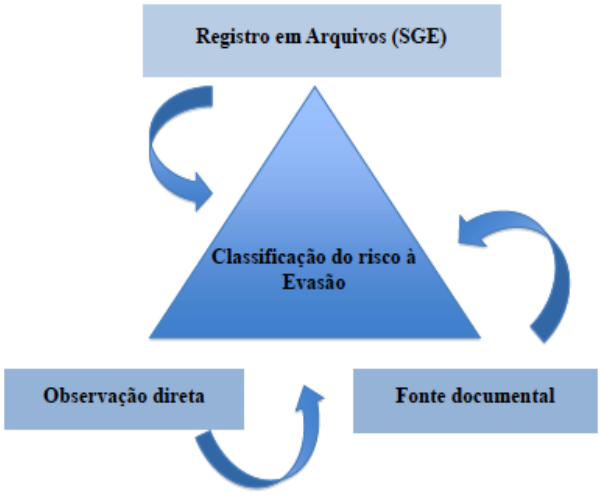

Figura 6: Triangulação dos dados da pesquisa. Fonte: [22]

Protocolos eletrônicos também podem ser utilizados como ferramentas de prevenção a evasão estudantil como abordado em [22] com o desenvolvimento de uma planilha eletrônica onde foram armazenadas as características de cada perfil como modelagem apropriada para, a partir das combinações dessas características, indicarem em qual classificação de risco o aluno se enquadra. Conforme mostrado na figura 6 a partir da análise dos dados é construído o perfil do aluno propenso a evadir e a classificação do risco à evasão escolar em baixo, moderado baixo, moderado alto e alto.

Embora tenham sido apresentados alguns projetos que visam contribuir com a predição da evasão discente através da utilização de ferramentas que auxiliem a tomada de decisão dos profissionais de educação, tais projetos não têm como um dos principais objetivos, evitar que a análise destes dados seja comprometida pelo fator viés onde há uma diferença entre a predição média do modelo preditivo e o valor correto esperado. O projeto SDBayes, por exemplo, utiliza-se de poucas variáveis para calcular a propensão de evasão do discente desconsiderando dessa forma fatores que podem ocupar um peso maior na decisão do aluno de evadir determinado curso. Dessa forma a predição pode se tornar tendenciosa pela limitação dos possíveis cenários que poderiam afetar o desempenho do aluno.

Os Protocolos eletrônicos, por sua vez, apresentam uma preocupação a respeito de como os dados obtidos para análise são manipulados em relação à ética e sigilo das informações, mas possuem a mesma limitação que o SDBayes em relação à quantidade de variáveis ponderadas para calcular a propensão de evasão do aluno. Além disso, o aplicativo foi desenvolvido para dispositivos android nativos, o que impossibilita o suporte a multiplataforma e o limita aos recursos de hardware disponíveis no aparelho utilizado.

Neste contexto, o presente trabalho ateve-se a construir um produto minimamente viável para auxiliar na tomada de decisão quanto a evasão escolar, baseando a definição do conjunto de fatores e a não-interferência do viés nos resultados preditivos na experiência da equipe pedagógica multidisciplinar do IFBA/Campus Porto Seguro.

IV. Metodologia

Para o presente trabalho não utilizaremos um algorítimo de automatização, mas sim, a análise visual dos dados de alunos que cursavam o ensino médio integrado, visando gerar conhecimentos para profissionais da área pedagógica, dirigidos à solução de um problema real. Quanto a abordagem, o trabalho apresenta um caráter quantitativo enumerando de forma probabilística a potencialidade de um aluno evadir o curso, através da análise de fatores que possibilitam a antecipação da tomada de decisão com o objetivo de evitar uma possível evasão escolar.

O trabalho teve como primeira etapa o diálogo com especialistas na área pedagógica e de acompanhamento estudantil. Em seguida, a identificação e extração dos dados relevantes para o projeto, a leitura de documentação referente a atuações semelhantes e seus resultados, bem como estruturar a lógica que indicaria a maior probabilidade de evasão. Em seguida, foi desenvolvido um protótipo que, ao final, foi entregue a

especialistas, que lidam direta e indiretamente com o problema, no intuito de obter resultados reais.

V. Solução Desenvolvida

Nesta seção, será apresentado o passo a passo na construção do Monitum que, etimologicamente, vem do latim e significa aquele que adverte, que dá conselho, que faz pensar. Esse sistema busca orientar as ações da equipe pedagógica acerca da evasão discente no Instituto Federal da Bahia — Campus Porto Seguro, considerando os fatores mais verificados no campus. A primeira etapa mostrará a extração, exploração, preparo e operacionalização dos dados de alunos que cursaram o 1º ano dos cursos técnicos em Alimentos, Biocombustíveis e Informática no ano de 2019, dados oriundos da pesquisa sobre evasão escolar realizada [5].

Por fim, a segunda etapa mostrará a construção de um sistema que retorna, em ordem decrescente, os alunos que tiverem uma maior pontuação, dada a existência dos fatores considerados causadores da maior parte da evasão do campus. Nele, o usuário poderá inserir importantes informações acerca do perfil do aluno avaliado:

- Matrícula
- Ano letivo
- Série em curso
- Relação dos fatores
- Observações

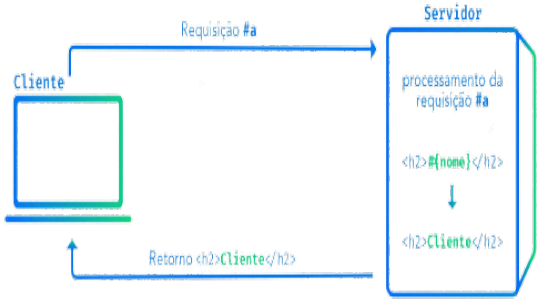

Figura 7: Arquitetura da solução proposta para o Monitum. Fonte: [23]

Portanto, como demonstrado a figura 7, temos uma solução que através de requisições HTTP, acessa uma base de dados construída com PostgreSQL [24] e PHP (*Hypertext Preprocessor*) [25]. Essa base de dados é composta por dados cedidos pelos especialistas da área pedagógica do Campus Porto Seguro.

A. *Pipeline para a Análise dos Dados*

Como já mencionado na Fundamentação Teórica, na subseção Análise de Dados, todo e qualquer trabalho que envolva a análise de dados consiste em, pelo menos, seis fases: Extração, Preparação, Planejamento do Modelo, Construção do Modelo, Comunicação dos Resultados e Operacionalização [16]. Para a construção da primeira parte do sistema Monitum, seguiu-se o fluxo como mostra a figura 8, onde cada etapa será elucidada a seguir.

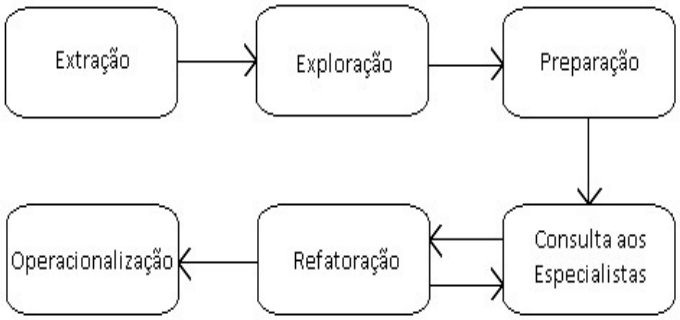

Figura 8: Fluxo do processo de Análise de Dados adotado para este trabalho. Fonte: Autor

B. *Extração*

A fase de extração consiste na obtenção de dados originados de diversas fontes: aplicativos mobile, sensores, páginas web ou até mesmo dados estruturados, semiestruturados e não estruturados, oriundos de bases convencionais, suportadas ou não por SGBD [13]. No caso deste trabalho, os dados coletados foram cedidos pela equipe pedagógica multidisciplinar do IF-BA/Campus de Porto Seguro. Esses são dados não-estruturados obtidos através de planilhas no formato XLS como mostra a figura 9.

C. *Exploração*

Uma vez coletados os conjuntos de dados necessários ao projeto, fez-se necessária a familiarização com os mesmos. Nesta fase, o objetivo foi conhecer o máximo possível dos dados para identificar ajustes a serem realizados na próxima fase do fluxo — preparação. O projeto, portanto, trabalhou

	A	B	C	D	E	F	G	H
1	**Aluno**	Aluno 1	Aluno 2	Aluno 3	Aluno 4	Aluno 5	Aluno 6	Aluno 7
2	**Turma**	1º TA A	1º TA A	1º TA A	1º TA A	1º TA A	1º TA A	1º TA A
3	**Nº de Matrícula**	**********	**********	**********	**********	**********	**********	**********
4	Faltas recorrentes				x			
5	Dificuldades no deslocamento para a escola (falta de transporte, distância entre a residência e a escola etc)							
6	Problemas comportamentais				x			
7	Problemas cognitivos							
8	Desinteresse pela área/curso	x						
9	Defasagem educacional		x	x				X
10	Dificuldades em se relacionar com os colegas							
11	Reprovado em anos letivos nas escolas anteriores	x	x					
12	Repetente no IFBA		x	x				
13	Vítima de bullying							
14	Distorção idade/série	x				x		x
15	Cotista	x	x	x	x	x		

Figura 9: Planilha utilizada para consulta dos dados dos alunos. Fonte: Autor

com dados de alunos que foram reprovados, aprovados e que evadiram.

- **Matrícula** — Contém o identificador do aluno associado ao ano de ingresso na instituição.
- **Ano letivo** — Refere-se ao período do ano no qual são desenvolvidas as atividades escolares efetivas.
- **Série em curso** — São numerados do 1º a 4º ano no contexto do ensino médio integrado.

D. Preparação

Na fase de preparação foram aplicados alguns ajustes nos dados buscando melhorar sua utilização para análise. Aqui, os valores ausentes e as variáveis irrelevantes ao projeto não foram consideradas nas tabelas. Além disso, foram feitas filtragens para ordenar os registros e definidas as relações associativas entre as tabelas como mostra a figura 10.

1) AlunoFator: A tabela alunoFator é resultado da junção dos dados das tabelas aluno e fator. O tipo da coluna *resposta* foi alterada para *integer* devido às comparações feitas no cálculo do potencial de evasão.
2) Aluno: Foi criada a variável observação para que ao cadastrar um aluno fossem adicionados comentários ao registro do mesmo, para que essa informação pudesse ser consultada posteriormente pelo profissional da área pedagógica.
3) Ranking: No ranking, a pontuação é resultado da soma dos pesos de cada fator que foi associado positivamente ao aluno. Essa variável consiste em um valor que é calculado considerando os pesos definidos pelos especialistas e é atribuído a cada aluno.

E. Consulta aos Especialistas

Nesta fase foi consultada a profissional da equipe pedagógica multidisciplinar do IFBA/Campus Porto Seguro que atua diretamente com os alunos e que também é a autora da pesquisa sobre evasão escolar da qual se originou o desenvolvimento do Monitum [5]. A pesquisadora informou que os critérios utilizados como referência para a elaboração do Sistema Monitum foram levantados em pesquisa realizada pela Secretaria de Educação Profissional e Tecnológica (MEC/SETEC).

Durante a entrevista foi feita uma contextualização do Monitum e discutido o funcionamento da ferramenta e como os dados dos alunos estavam sendo estruturados. Terminada essa pequena introdução, foi discutido como as variáveis seriam utilizadas na soma da pontuação do potencial de evasão dos alunos.

Para Góis [5], visto que a evasão escolar é um processo dinâmico, influenciado por um conjunto de condições que não ocorrem isoladamente, presume-se que no decorrer do seu percurso escolar alguns estudantes já revelam indícios quanto à possibilidade de abandono escolar. É nesse cenário que o Sistema Monitum busca atuar, monitorando e apontando precocemente essas fragilidades demonstradas pelos alunos, de modo que as ações e decisões tomadas sejam compatíveis com a realidade do educando e em prol de sua permanência e êxito na instituição.

F. Refatoração

Como mencionado na subseção Preparação, presente na seção Solução Desenvolvida, foi incluída à tabela Aluno uma

nova variável chamada OBSERVACAO, responsável por conter informações específicas do perfil do aluno avaliado.

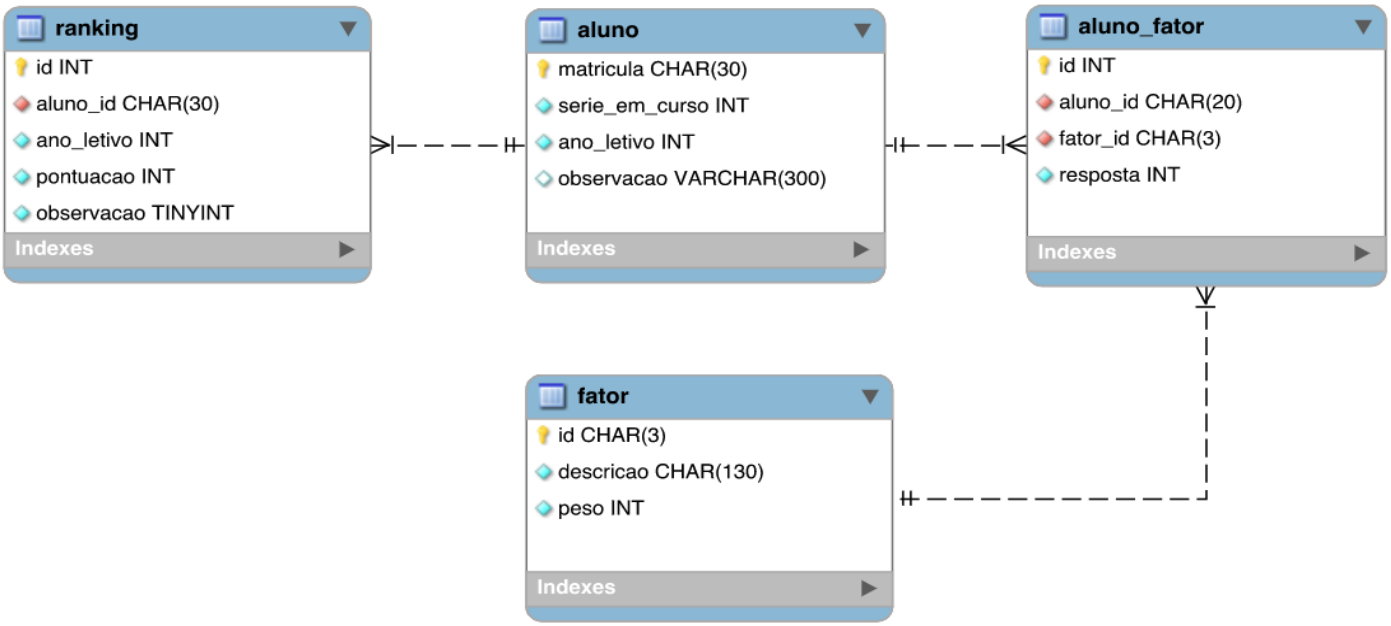

Figura 10: Diagrama utilizado na construção da base de dados.

Com a finalização da entrevista, voltou-se a etapa Preparação para que fosse possível melhorar o projeto com a inclusão das modificações que os profissionais também julgaram ser de suma importância.

G. Operacionalização

Na última fase do processo de desenvolvimento do Monitum, foi criado um mínimo produto viável para apresentar a listagem da pontuação dos alunos e ordená-los decrescentemente. O mínimo produto viável consiste em um sistema web criado com a linguagem de programação PHP. Abaixo, as informações referentes as tecnologias utilizadas e as funcionalidades do sistema serão melhor detalhadas.

1) *Tecnologias Utilizadas:* Inicialmente o Monitum foi pensado para ser um sistema web *server-side* [23] — solução utilizada para tranformar páginas web estáticas e sem funcionalidade em páginas dinamicas, tendo como linguagem de programação o PHP para o desenvolvimento de aplicações presentes e atuantes no lado do servidor.

Feitas as considerações sobre a arquitetura do sistema, escolheram-se as tecnologias que seriam utilizadas na construção da mesma.

- PHP — Criado em 1995 por Rasmus Lerdorf é uma linguagem de programação interpretada livre, que possui mais de 20 anos de existência sendo uma das mais utilizadas no mundo, com um vasto conjunto de APIs (*Application Programming Interface*).
- PostgreSQL — É um sistema de banco de dados relacional de código aberto com mais de 30 anos de desenvolvimento possui uma estrutura confiável e robusta de recursos e desempenho. Possui uma comunidade ativa que oferece muitos lugares úteis para se familiarizar com a tecnologia.

2) *Funcionalidade Cadastro:* Inicialmente são inseridas as informações necessárias para criar a massa de dados utilizada para calcular a pontuação de cada aluno. Para que o sistema efetue o cadastro são solicitadas as seguintes informações:
   - Matrícula
   - Ano letivo
   - Série em curso
   - Questionário de fatores

   O questionário de fatores é composto por 30 perguntas definidas pela equipe pedagógica multidisciplinar do Campus, fruto da análise do perfil acadêmico e socioeconômico dos alunos e do Documento Orientador [26] elaborado pela SETEC/MEC, onde foram mapeados quais daquelas razões correspondiam à realidade dos alunos como mostra a tabela 1.

3) *Funcionalidade Pesquisa:* A partir da matrícula do aluno cadastrado é possível realizar um filtro na consulta para obter a pontuação, os dados cadastrados no questionário de fatores como também informações pessoais do aluno.

4) *Funcionalidade Ranking:* O sistema retorna uma lista ordenada decrescentemente baseada no cálculo da

pontuação do aluno. Esse valor é calculado considerando os pesos que foram definidos pelos especialistas dos fatores que foram respondidos positivamente no questionário de cadastro. Segue a tabela com os fatores específicos que aumentam as chances de Evasão e de Retenção, de acordo com o Documento Orientador para a Superação da Evasão e Retenção na Rede Federal de Educação Profissional, Científica e Tecnológica [26].

Tabela I: Fatores de Vunerabilidade Escolar.

FATORES DE VULNERABILIDADE ESCOLAR		PESO (01 a 05)
1	Faltas recorrentes	5
2	Dificuldades no deslocamento para a escola (falta de transporte, distância entre a residência e a escola etc)	2
3	Problemas comportamentais	1
4	Problemas cognitivos	4
5	Desinteresse pela área/curso	4
6	Defasagem educacional	5
7	Dificuldades em se relacionar com os colegas	4
8	Reprovado em anos letivos nas escolas anteriores	3
9	Repetente no IFBA	4
10	Vítima de bullying	4
11	Distorção idade/série	4
12	Cotista	2
13	Reprovado em duas unidades na mesma disciplina no ano letivo corrente	3
14	Indisponibilidade de tempo para participar de atividades no contra turno	3
15	Estudante já concluiu o ensino médio em outra instituição	2
16	Vulnerabilidade socioeconômica	3
17	Tem filhos	4
18	Falta de acompanhamento da família no processo ensino/aprendizagem	3
19	Trabalha no turno oposto às aulas	5
20	Estudante não reside com o tutor legal	2
21	Problemas psicológicos	4
22	Conflitos familiares	3
23	Estudante tem acompanhamento da rede de proteção local (Conselho Tutelar/CRAS/CREAS/CAPS)	3
24	Estudante envolvido com drogas/atos infracionais	4
25	Problema de saúde do estudante	4
26	Ambiente doméstico desfavorável ao estudo	3
27	Cuidador de parentes no turno oposto (idoso, criança, pessoa com necessidade especial)	2
28	Estudante não tem rotina de estudos	4
29	Divergência e/ou intolerância religiosa	3
30	Estudante com Necessidades Educativas Especiais	3

O Monitum é um protótipo de modelo de classificação, do tipo SAD (*Sistema de Apoio à Decisão*) que busca indicar a propensão de evasão de aluno a partir das informações contidas numa base de dados, oriunda de um SGBD, em uma base de conhecimento probabilística [27]. Este é um modelo que está em transformação a partir da experiência dos profissionais da área pedagógica. Abaixo as figuras 11, 12 e 13 demonstram a interface web criada para a realização do cadastro dos alunos e a visualização das suas respectivas classificações.

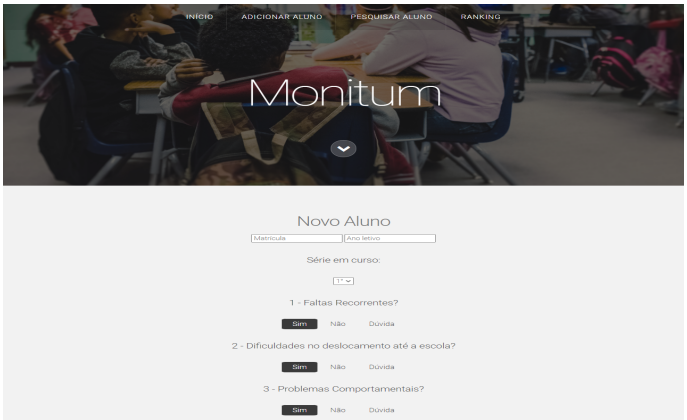

Figura 11: Módulo - Cadastrar Aluno

Figura 12: Módulo - Pesquisar Aluno

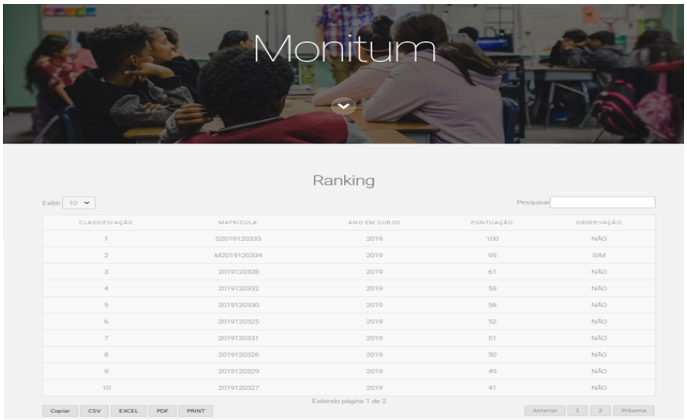

Figura 13: Módulo - Ranking

VI. Resultados Obtidos

A pesquisa utilizou na base de dados um universo de 180 alunos matriculados no 1º ano dos cursos integrados ao ensino médio, com faixa etária compreendida entre os 14 e 17 anos. O período de coleta dos dados considerados para esse trabalho durou cerca de dois meses e obteve um resultado de respostas dos alunos próximo a 22,2%. É necessário ressaltar que o período pandêmico que se estende desde o final de 2019 afetou a obtenção dos dados dos alunos após as aulas presenciais terem sido suspensas, porém, a pesquisa foi com base nos resultados dos alunos após o término do ano letivo, ou seja, podemos destacar uma maior confiabilidade nas respostas obtidas.

Como resultado, foi possível obter diversos relacionamentos entre as características dos alunos, permitindo melhor conhecer o seu perfil e possibilitando a utilização das informações em projetos e ações futuras. Características como distorção de idade/série, cotista, tempo dedicado ao estudo, não residir com tutor legal e dificuldade no deslocamento para a escola, entre outras, poderão ser sobrepostas, permitindo associar possíveis influências destas sobre a questão da evasão.

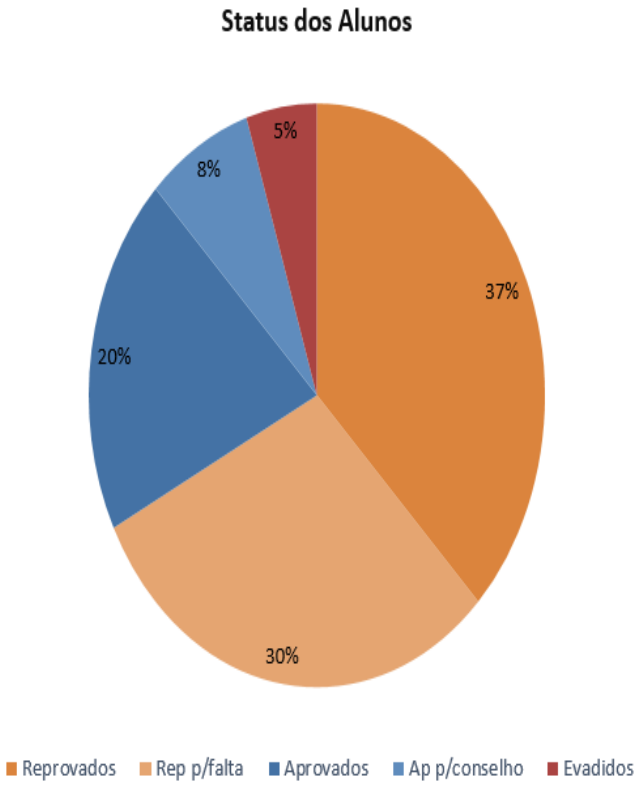

Figura 14: Status dos alunos avaliados.

A figura 14 demonstra, de forma percentual, de que maneira está avaliado o resultado de conclusão dos alunos matriculados nos cursos integrados ao ensino médio do IFBA Campus Porto Seguro. Ainda de acordo com a figura, somente 28% dentre os alunos pesquisados foram aprovados por atingirem o percentual mínimo de participação das aulas e por alcançarem a pontuação miníma da instituição de ensino. A influência dos fatores de risco no coeficiente de rendimento do aluno é perceptível, como demonstrado na figura 14. Nota-se, que os alunos que evadiram em seus respectivos cursos estão entre as maiores pontuações como demonstrado nas figuras 15, 16, 17 e 18.

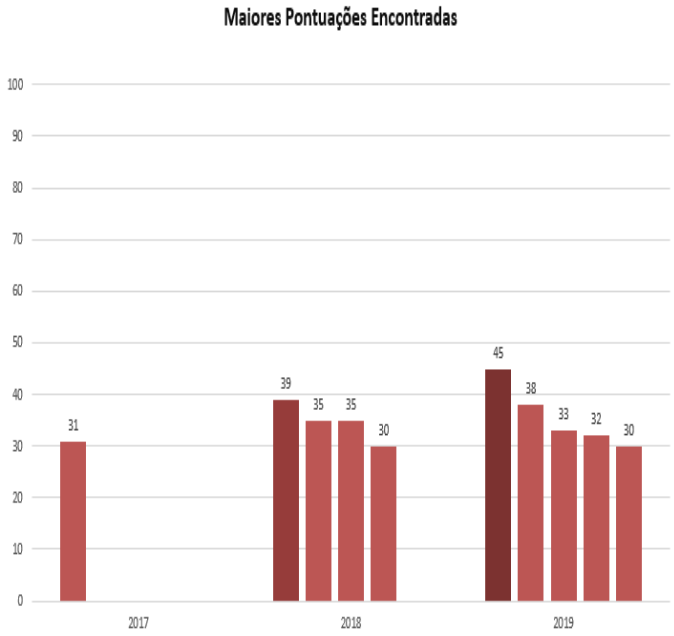

Figura 15: Maiores pontuações de evasão encontradas por ano de ingresso.

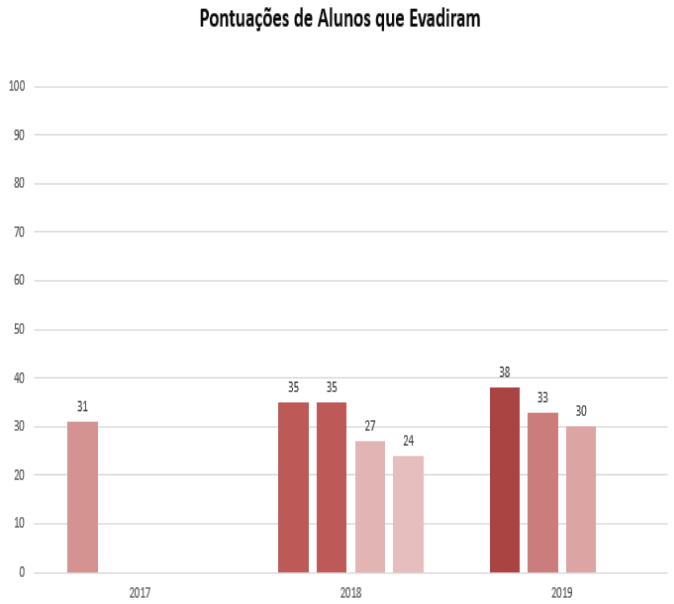

Figura 16: Pontuações de alunos que evadiram por ano de ingresso.

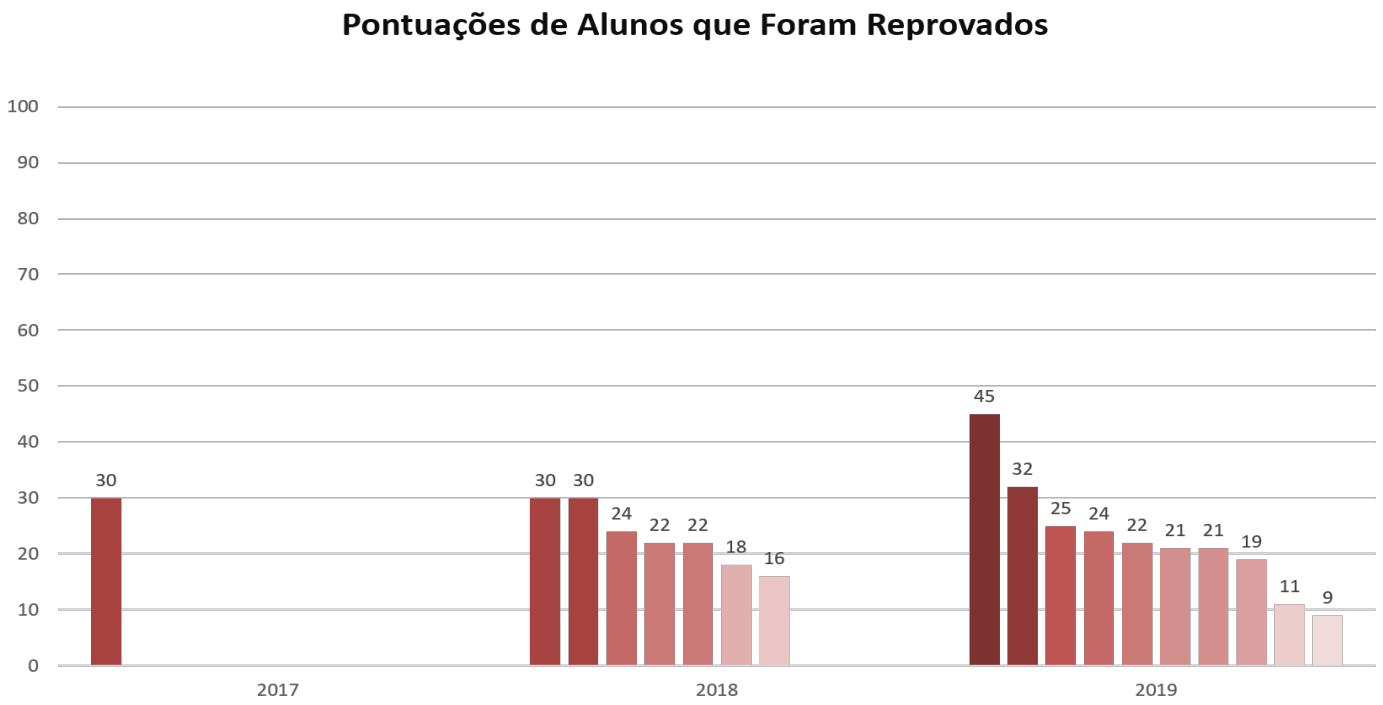

Figura 17: Pontuações de alunos que foram reprovados por ano de ingresso.

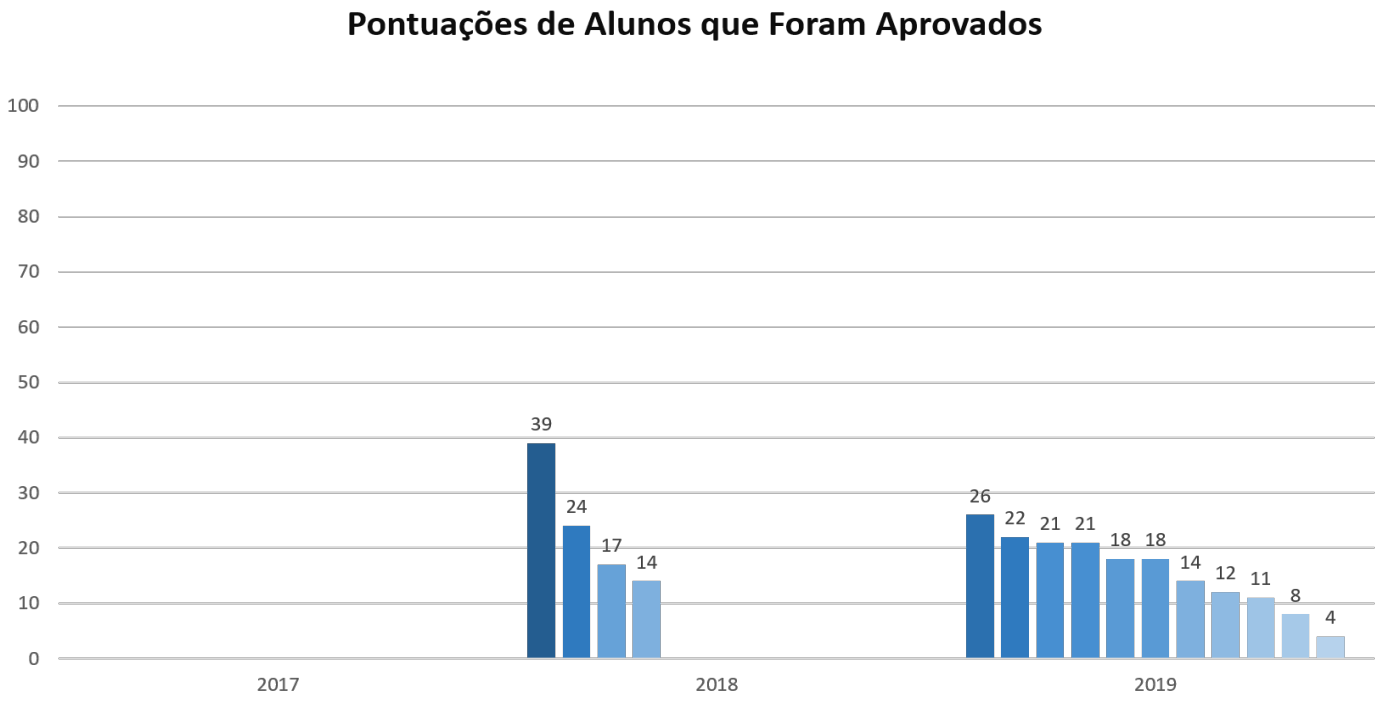

Figura 18: Pontuações de alunos que foram aprovados por ano de ingresso.

Os fatos apresentados pelos resultados nas figuras 19 a 21 trazem perdas pedagógicas significativas para o processo de ensino-aprendizagem deste grupo de alunos. Este fato termina esvaziando muitas das salas de aula, fazendo com que alguns professores terminem suas aulas mais cedo para não prejudicar a maioria dos alunos. Neste ponto, acreditamos que as medidas e intervenções pedagógicas que a equipe multidisciplinar do campus onde a pesquisa foi realizada, munida dessas informações, podem garantir a permanência do aluno na instituição antecipando potenciais caso de evasão.

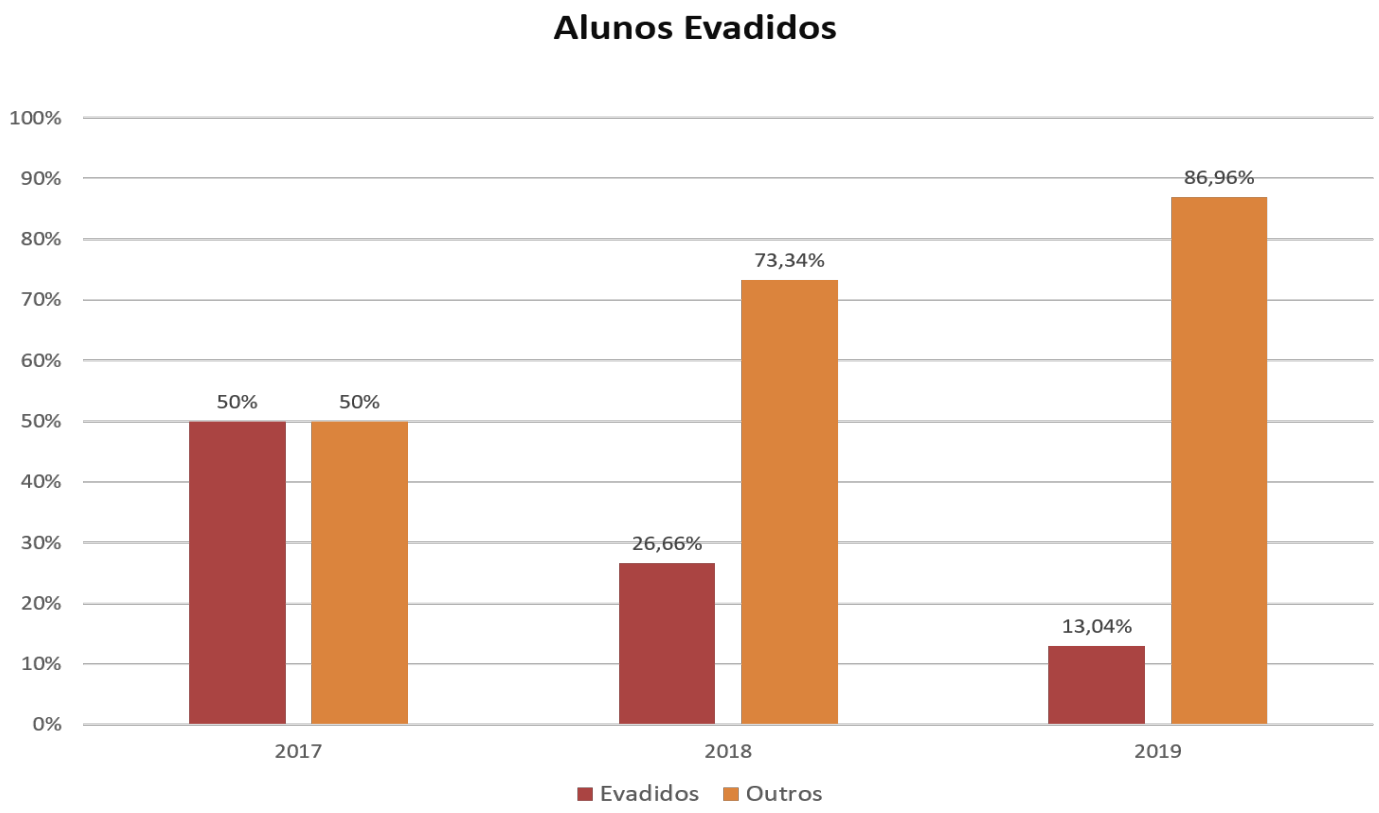

Figura 19: Alunos que evadiram ingressantes de 2017, 2018 e 2019.

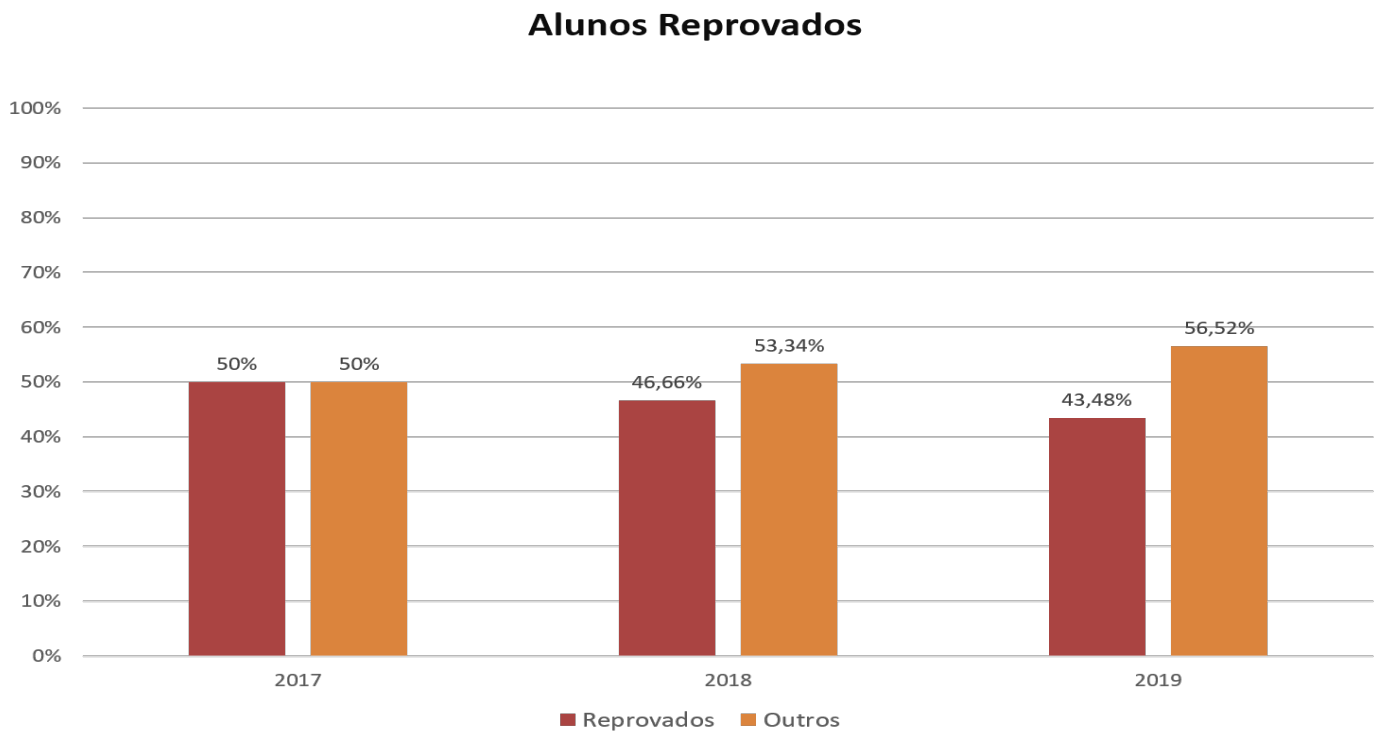

Figura 20: Alunos reprovados ingressantes de 2017, 2018 e 2019.

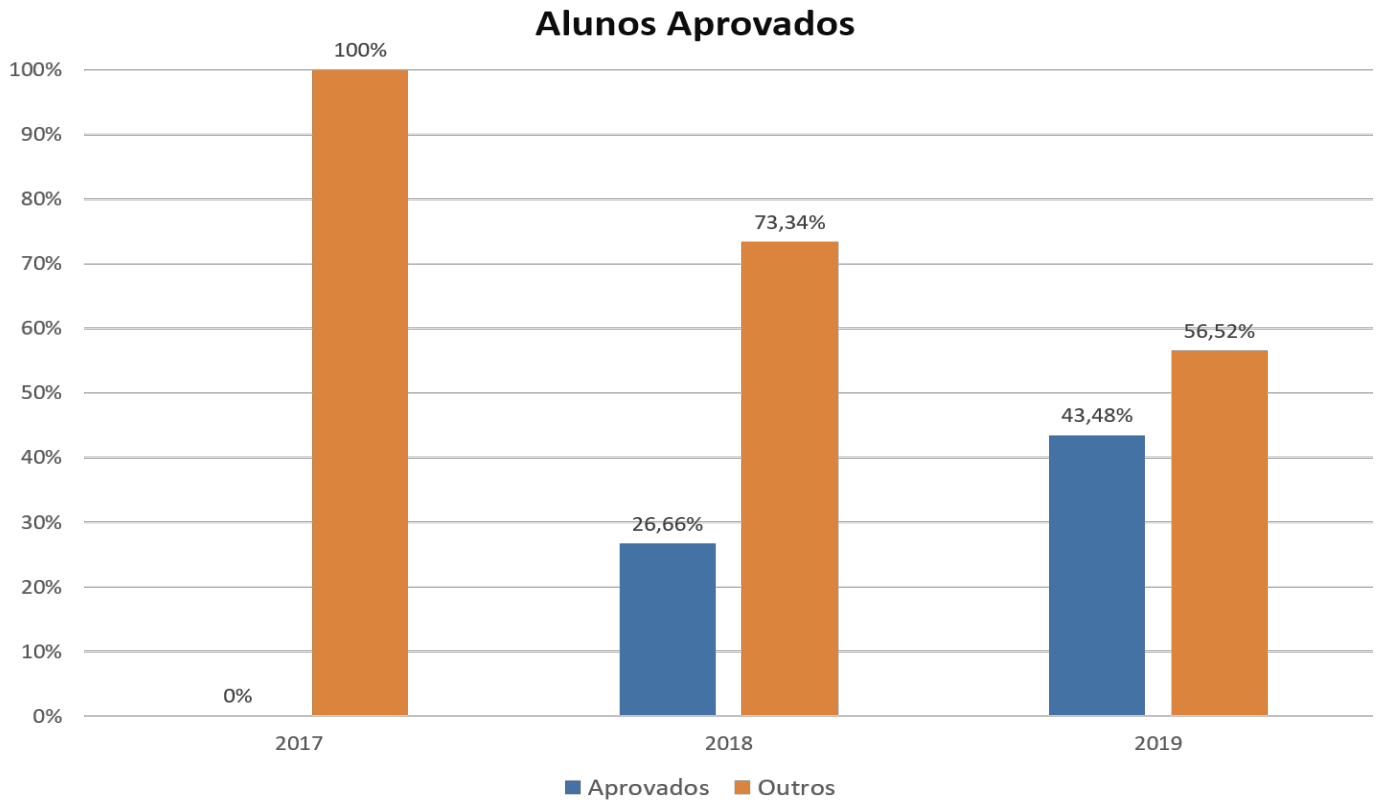

Figura 21: Alunos aprovados ingressantes de 2017, 2018 e 2019.

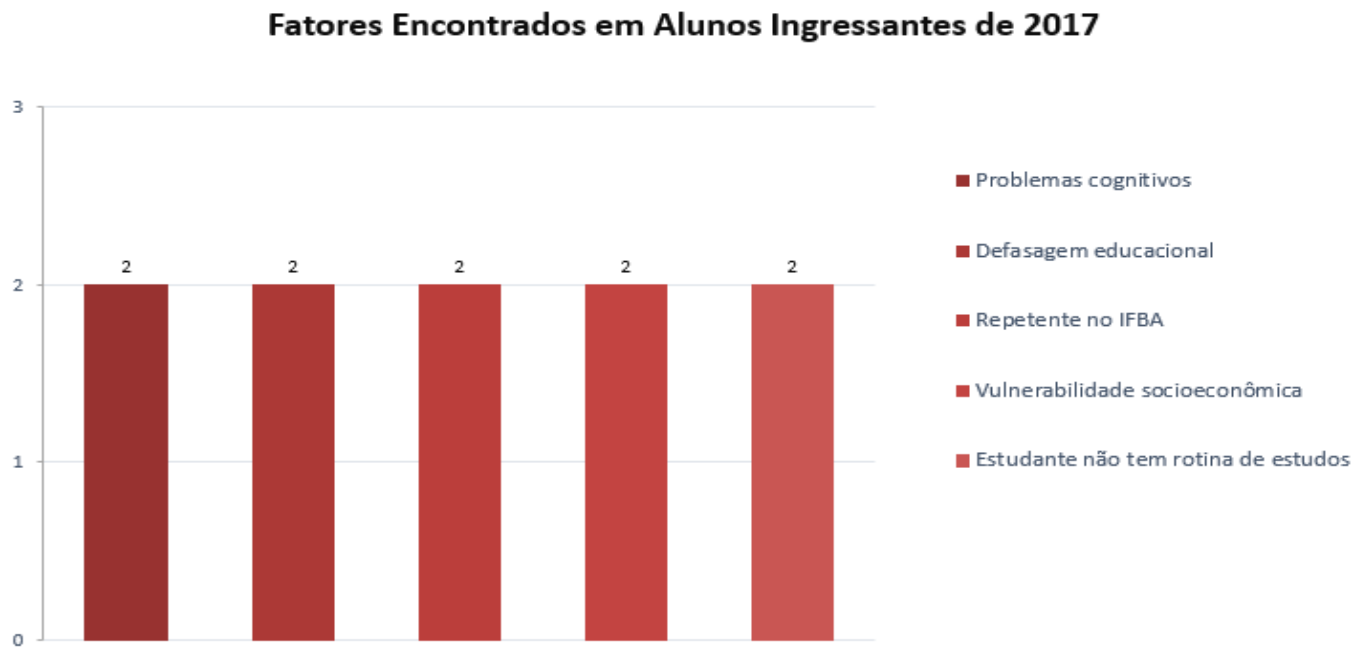

Figura 22: Quantidade de fatores encontrados em alunos ingressantes de 2017.

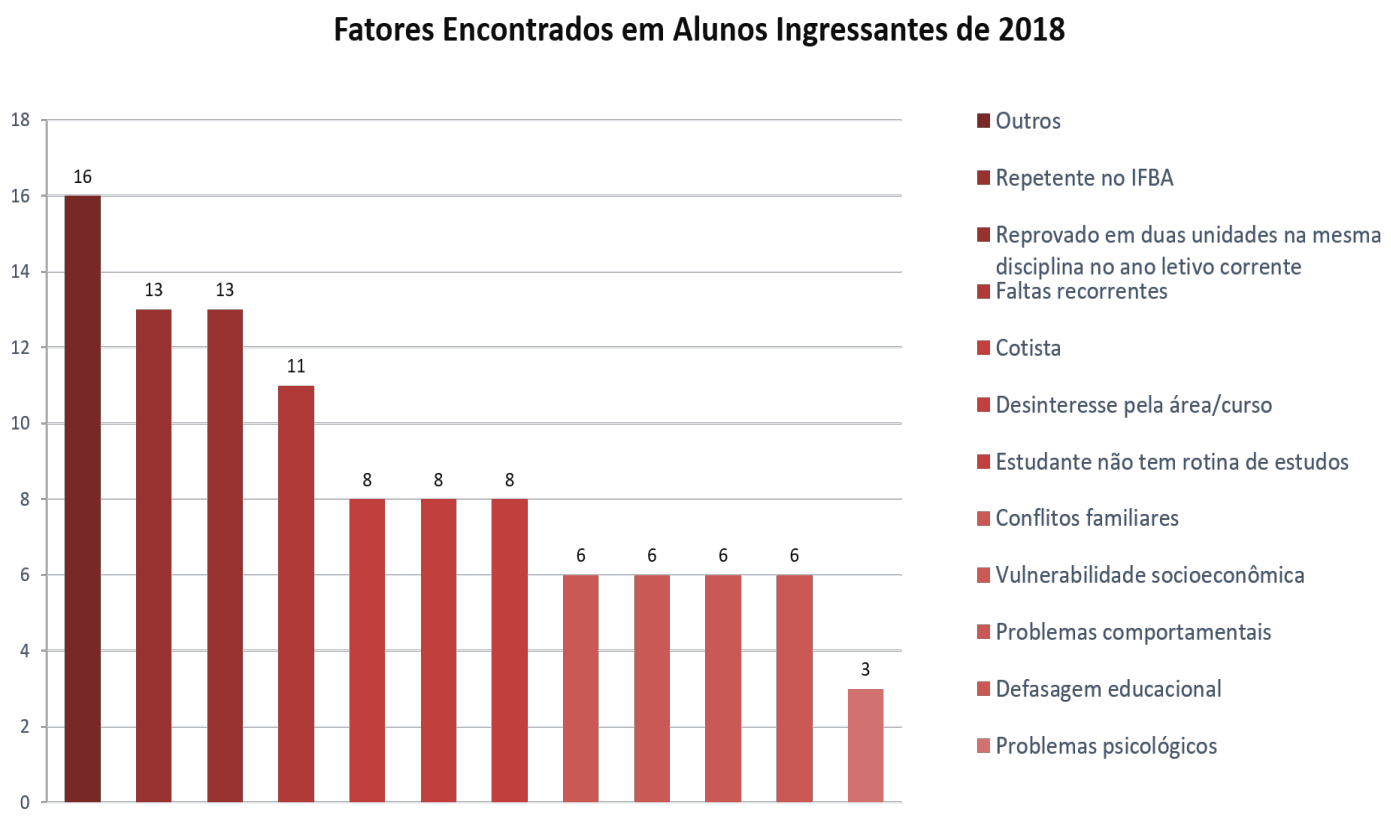

Figura 23: Quantidade de fatores encontrados em alunos ingressantes de 2018.

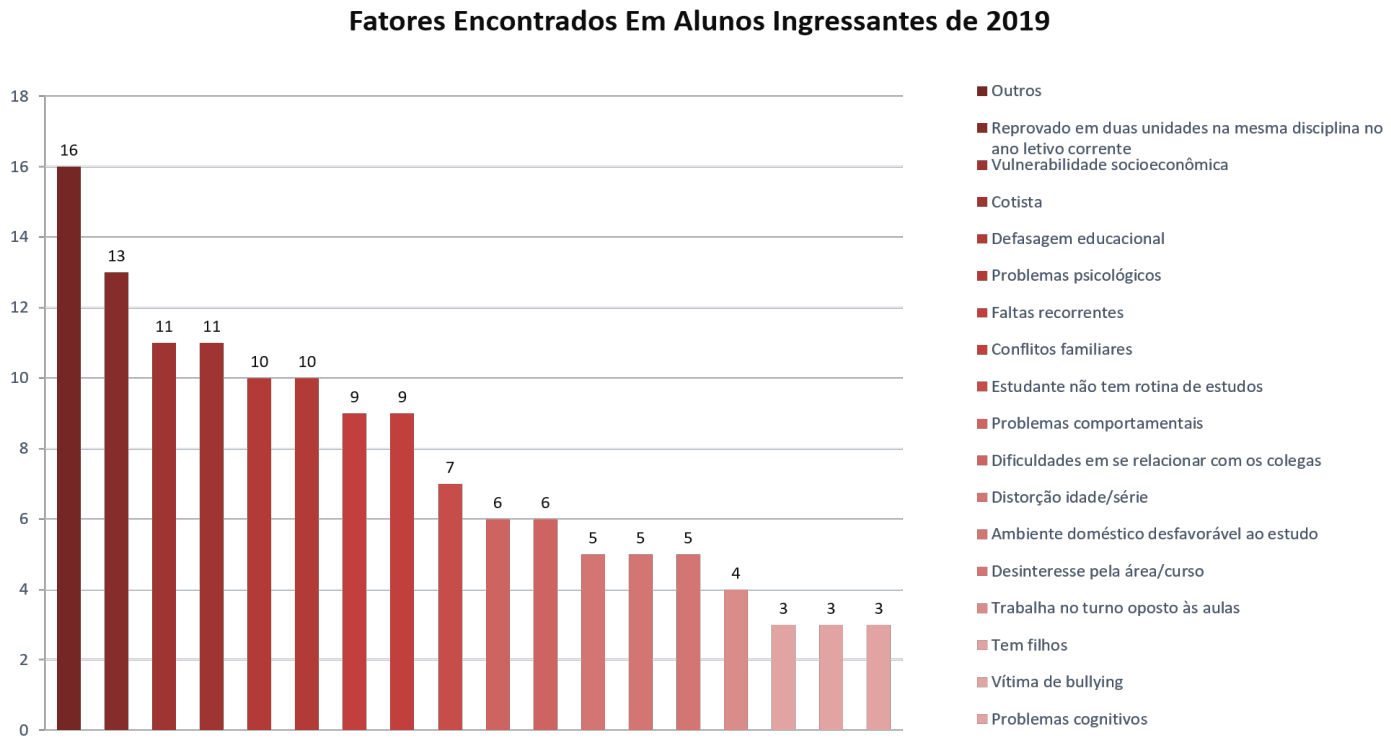

Figura 24: Quantidade de fatores encontrados em alunos ingressantes de 2019.

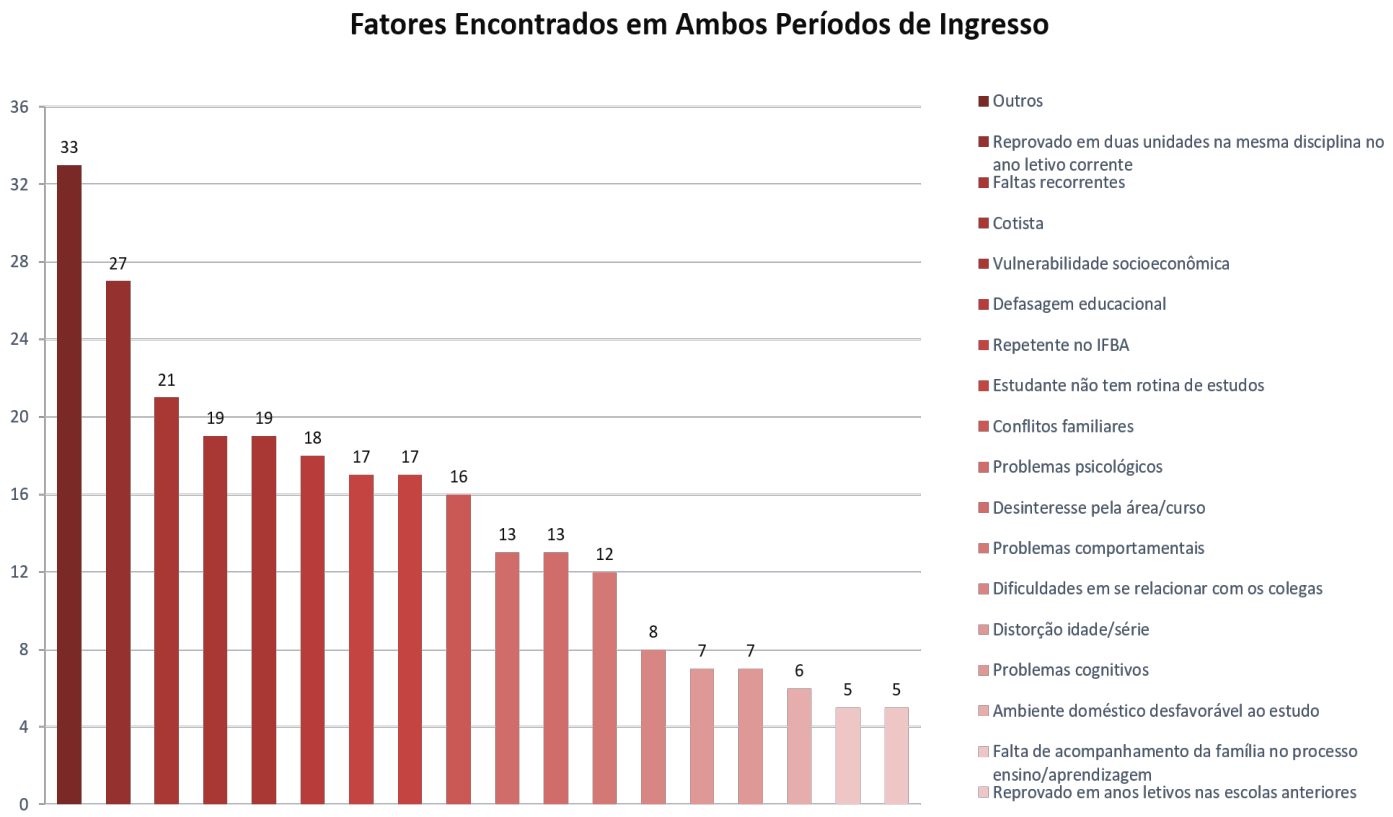

Figura 25: Quantidade de fatores encontrados em ambos períodos de ingresso.

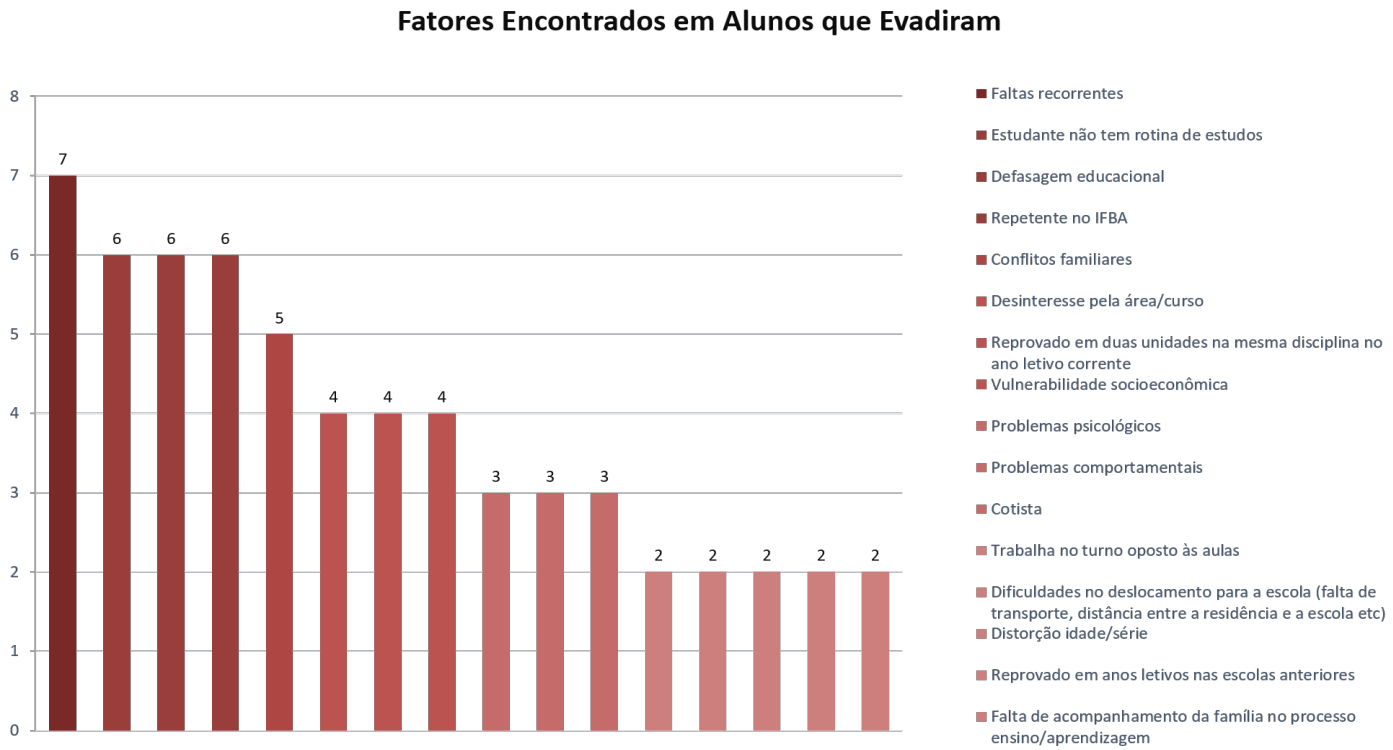

Figura 26: Quantidade de fatores encontrados em alunos que evadiram.

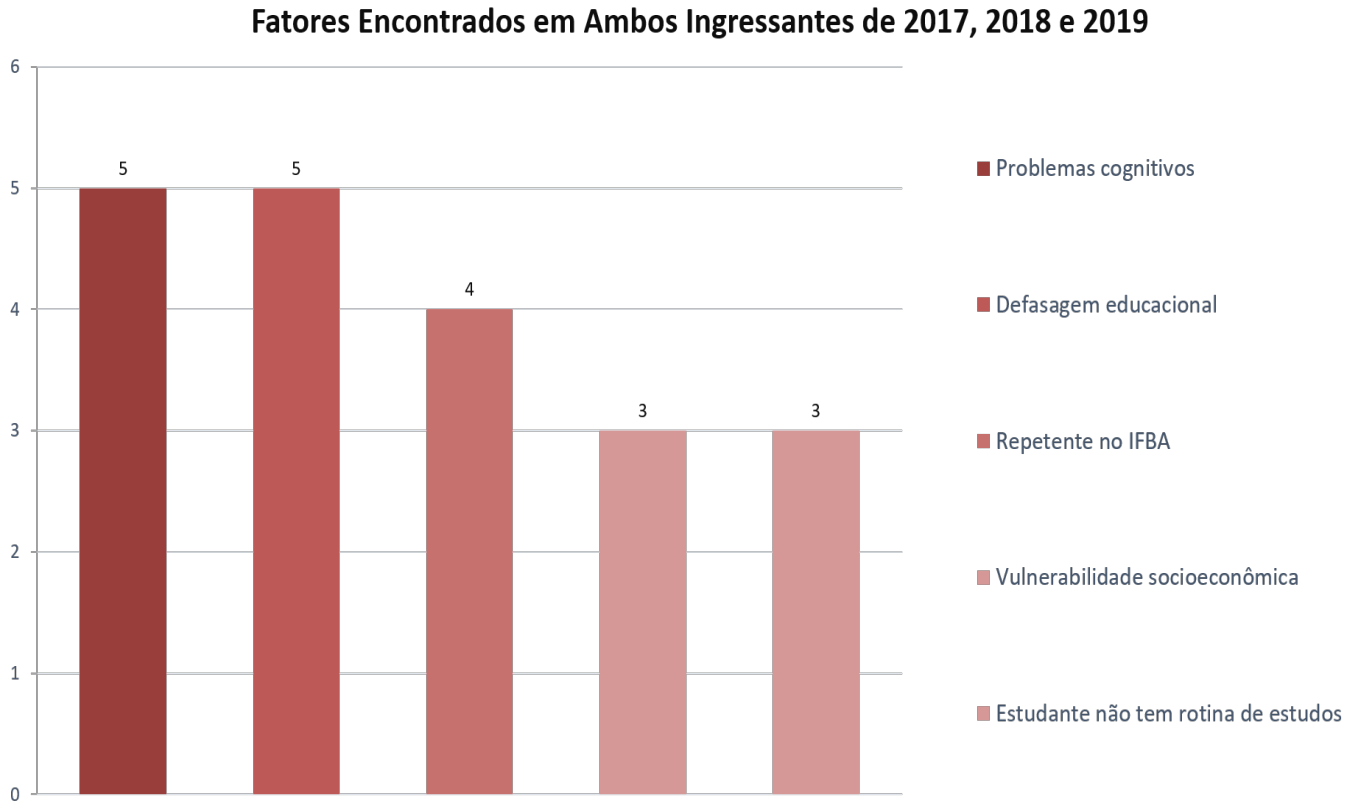

Figura 27: Quantidade de fatores encontrados em ambos ingressantes de 2017, 2018 e 2019.

Com relação à parte de vulnerabilidade dos alunos, conforme indicado na figura 22, os alunos ingressantes de 2017 apresentaram os mesmos fatores. A figura 23 mostra a quantidade dos fatores encontrados nos alunos ingressantes de 2018 onde dentre os fatores encontrados destacaram-se: *Repetente no IFBA* e *Reprovado em duas unidades na mesma disciplina no ano letivo corrente*. A figura 24 mostra que há uma maior variedade de fatores encontrados nos ingressantes de 2019 em relação aos ingressantes de 2017 e 2018. A figura 25 apresenta os fatores que se repetem em ambos ingressantes, destacando o fator *Reprovado em duas unidades na mesma disciplina no ano letivo corrente* como o mais notado nos perfis dos alunos avaliados.

Em relação à figura 26, nota-se que o fator *Faltas recorrentes* foi o mais associado aos alunos que evadiram em comparação ao fator *Reprovado em anos letivos nas escolas anteriores* que indica um histórico escolar negativo. Apenas 2 fatores foram encontrados estritamente nos ingressantes de 2018 demonstrado pela figura 27. Já os ingressantes de 2019 apresentaram mais fatores em contraste aos de 2018, porém com fatores que apontam vulnerabilidades sociais como o fato de sofrer *bullying* e intolerâncias religiosas. A figura 29 mostra os fatores de vulnerabilidade que se perpetuaram em ambos ingressantes. Ao iniciar as atividades escolares das turmas de 1º ano dos cursos integrados ao ensino médio são ofertadas duas turmas para cada curso, porém com o alto número de evasão e reprovação no segundo ano essas turmas são unificadas como consequência do baixo de número de alunos aprovados. Em geral, como análise final, os cursos Técnico em Biocombustíveis e Técnico em Informática tiveram o maior grau de evasão 37,5% respectivamente, em comparação ao curso Técnico em Alimentos com 25% do total de alunos evadidos.

VII. Conclusão

Segundo Gois [5], na pesquisa realizada para a implantação desse sistema, entre os anos de 2015 e 2018, mais de 60% dos alunos que evadiram no Campus Porto Seguro o fizeram enquanto cursavam o 1º ano. De acordo com o objetivo principal deste trabalho, foi desenvolvida uma ferramenta que auxilie os profissionais da área pedagógica na tomada de decisão, monitorando e apontando precocemente essas fragilidades demonstradas pelos alunos.

De acordo com Lencastre [8], o ano escolar em 2021 começou muito diferente do esperado. Havia a expectativa de um retorno gradual às aulas presenciais, mas o avanço da covid-19 em todo o país, a escassez de vacinas e, consequentemente, a lentidão do programa de vacinação estão mantendo parte das escolas fechadas. Duas pesquisas recentes mostram que o Brasil regrediu 20 anos nas taxas de abandono e de evasão escolar. Com a pandemia e sem uma coordenação nacional dos três níveis de governo, a desigualdade na educação brasileira fica ainda mais nítida.

Ainda que existam sistemas que trabalhem com a análise da evasão escolar, não foi possível identificar alguma solução

que atue de forma mais direta com os fatores reais vivenciados pelos alunos e que dê suporte a especialistas da área pedagógica a identificar precocemente o problema da mesma forma que este projeto se propõe trabalhar. Dos projetos citados da seção Trabalhos Correlatos, o projeto Protocolos Eletrônicos trabalha de forma mais direta do que o SDBayes ao apresentar mais variáveis para classificar o risco a evasão e ao considerar a variância de amostragem, mas ainda possui limitações quanto a quantidade e qualidade das variáveis selecionadas, afim de evitar o viés.

Esse trabalho surge nesse contexto como uma iniciativa necessária que visa auxiliar na tomada de decisão, proporcionando um modelo classificativo construído a partir da experiência dos profissionais da área pedagógica que acompanham as mudanças dos fatores que afetam o rendimento e permanência do aluno na instituição de ensino. O projeto não se limitou apenas a colher informações sobre a vida acadêmica dos alunos dos cursos integrados ao ensino médio, mas preocupou-se, também, em obter informações sociais e econômicas que possuem impactos significativos e que por sua vez culmina na baixa expectativa do aluno permanecer e concluir seus estudos.

Acredita-se assim que uma parcela significativa dos alunos que tenderiam a evadir seja acompanhada e amparada, visto que foi possível pensar em um projeto que futuramente poderá ser capaz de munir profissionais da área pedagógica a tomar ações preventivas de acordo com os resultados obtidos pelo uso da ferramenta Monitum. Com base nisso, foi construído um produto minimamente viável onde o usuário poderá inserir importantes informações acerca do aluno monitorado: **Matricula** — registro de ingresso na instituição; **Ano Letivo** — ano no qual são desenvolvidas as atividades escolares efetivas; **Série em Curso** — correspondente a uma etapa de ensino do ano letivo — e **Fatores de Vulnerabilidades** — informações coletadas sobre os alunos pela equipe pedagógica multidisciplinar do Campus. Com base nessas informações, uma lista com informações básicas dos alunos e seu potencial de evasão são mostradas em ordem decrescente.

A partir dos resultados obtidos, foi possível visualizar o cenário vivenciado pelos discentes, e utilizando-se do conhecimento das possíveis dificuldades que levam os alunos a evadirem ou a estenderem o tempo do curso, permitir que os profissionais da instituição tomem decisões baseadas em informações estratégicas.

VIII. Trabalhos Futuros

Durante o desenvolvimento da solução proposta, observaram-se possibilidades de melhorias na implementação visando evoluir o Monitum. Nesse contexto, destacam-se como trabalhos futuros:

- Construção da aplicação mobile de forma que seja acessível e prática.
- Possibilidade de importar os dados dos alunos de uma planilha.
- Disponibilização da funcionalidade onde o usuário possa filtrar a classificação dos alunos com base no ano letivo.
- Possibilidade de exibir gráficos estatísticos com base no ano letivo filtrado.

Referências

[1] E. C. da Silva Ferreira and N. M. de Oliveira, "Evasão escolar no ensino médio: Causas e consequências," *Scientia Generalis*, pp. 39–48, 2020, Centro Universitário de Patos de Minas.

[2] A. Neto, M. Modesto *et al.*, "Uma proposta de análise da evasão no ensino superior do ifba," *Ensino em Foco*, p. 3, 2018, Trabalho de Conclusão de Curso. Instituto Federal da Bahia (IFBA).

[3] R. L. L. E. S. Filho, P. R. Motejunas, O. Hipólito, and M. B. D. C. M. Lobo, "A evasão no ensino superior brasileiro," *Cadernos de Pesquisa*, pp. 641–659, 2007, Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia.

[4] G. Rosa, I. Pimentel, F. Sobrinho, and D. Rodrigues, "Mineração de dados no combate a evasão," *IV Workshop Nacional de Educação Profissional e Evasão Escolar*, 2018, V Colóquio Internacional sobre Educação Profissional e Evasão Escolar.

[5] L. S. Góis, "A Repercussão do Programa de Assistência e Apoio aos Estudantes na Evasão Escolar do Instituto Federal da Bahia - Campus Porto Seguro," *Programa de Pós-graduação em Educação Profissional e Tecnológica - IFBA*, 2020, Dissertação de Mestrado.

[6] R. Dore and A. Z. Lüscher, "Permanência e evasão na educação técnica de nível médio em minas gerais," *Docero*, pp. 772–789, 2011.

[7] I. C. e. C. UNICEF Brasil, "Enfrentamento da cultura do fracasso escolar," https://www.unicef.org/brazil/relatorios/enfrentamento-da-cultura-do-fracasso-escolar, 2021 (acessado em 22 de Novembro, 2021).

[8] C. Lencastre, "'pandemia' de abandono e evasão escolar," https://projetocolabora.com.br/ods4/pandemia-de-abandono-e-evasao-escolar/, 2021 (acessado em 22 de Novembro, 2021).

[9] D. M. Cunha, F. Fidalgo *et al.*, *Formação/Profissionalização de Professores e Formação Profissional e Tecnológica*, 2013, ISBN - 978-85-8229-010-1.

[10] G. Morrow, "Standardizing practice in the analysis of school dropouts," *Semantic Scholar*, pp. 342–355., 1986, Teachers College Record, v. 87, n. 3.

[11] H. Hamilton, "Overview of the kdd process," http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html, 2018 (acessado em 09 de Outubro, 2021).

[12] S. da Costa Côrtes, R. M. Porcaro, and S. Lifschitz, "Mineração de dados – funcionalidades, técnicas e abordagens," 2002.

[13] E. Navathe, *Sistemas de banco de dados. ISBN - 978-85-7936-085-5*. Pcarson Education, 2011.

[14] L. Digiampietri, F. Nakano, and M. Lauretto, "Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso," *Revista De Graduação USP*, pp. 17–23, 2016, Universidade de São Paulo. Escola de Artes, Ciências e Humanidades.

[15] J. W. Tukey, "The future of data analysis," *The annals of mathematical statistics*, pp. 1–67, 1962, v. 33, n. 1.

[16] E. E. Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Wiley, 2015, ISBN - 978-1118876138, n. 1.

[17] L. M. Carraro, "Modelos de predição | introdução à predição," https://medium.com/turing-talks/turing-talks-10-introdução-à-predição-a75cd61c268d, 2019 (acessado em 17 de Dezembro , 2020).

[18] B. Vasconcelos, "Poder preditivo de métodos de machine learning com processos de seleção de variáveis: uma aplicação às projeções de produto de países," 2017, Dissertação de mestrado. Universidade de Brasília.

[19] P. Tubella, "Estudantes britânicos derrotam o algoritmo de johnson," https://elpais.com/internacional/2020-08-18/los-estudiantes-britanicos-propician-la-primera-gran-derrota-de-johnson.html, 2020 (acessado em 17 de Dezembro, 2020).

[20] L. Fernando and P. V. Florentino, "Análise Multidimensional do Perfil Discente e Evasão no Ensino Superior. Trabalho de conclusão de curso. Instituto Federal da Bahia (IFBA)." 2013.

[21] J. Schreiber, A. Beskow, E. Nara, J. I. da Silva, J. de Moraes, and V. M. Najdzion, "Software: Sdbayes um auxílio para a predição de evasão discente," https://repositorio.ufsc.br/handle/123456789/181197, 2017 (acessado em 27 de Dezembro, 2021).

[22] I. D. S. Lima, “Classificação do risco por meio de protocolo eletrônico: Um instrumento para diagnóstico da evasão escolar — Dissertação de Mestrado. Programa de Pós-Graduação em Gestão e Tecnologias Aplicadas à Educação. Universidade do Estado da Bahia (UNEB).” 2017.

[23] A. H. Schneider, “Schneider, adolfo henrique. desenvolvimento web com client side rendering: combinando single page application e serviços de backend,” *Lume - Repositório Digital*, pp. 17–20, 2016.

[24] A. Milani, *PostgreSQL - Guia do Programador. ISBN - 978-8575221570*. Novatec Editora, 2008.

[25] T. Converse and J. Park, *PHP A Bíblia. ISBN - 978-8535211306*. Elsevier, 2003.

[26] Ministério da Educação, “Documento orientador para a superação da evasão e retenção na rede federal de educação federal de educação profissional, científica e tecnológica,” http://r1.ufrrj.br/ctur/wp-content/uploads/2017/03/Documento-Orientador-SETEC.pdf, 2014 (acessado em 20 de Dezembro, 2021).

[27] E. R. Neto, “E-bayes - sistema especialista para a análise da evasão discente de cursos de graduação no ensino superior,” https://repositorio.ufsc.br/handle/123456789/84446, 2002 (acessado em 27 de Dezembro, 2021).

IX. Apêndices

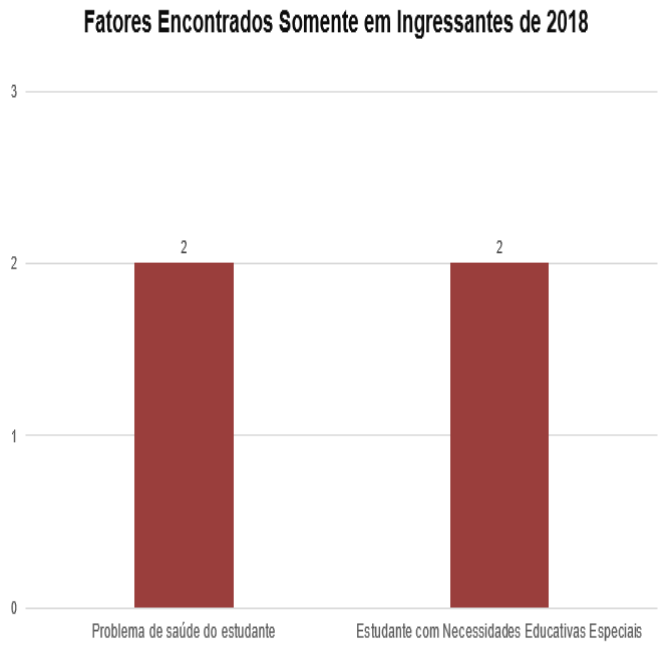

Figura 28: Quantidade de fatores encontrados somente em ingressos de 2018.

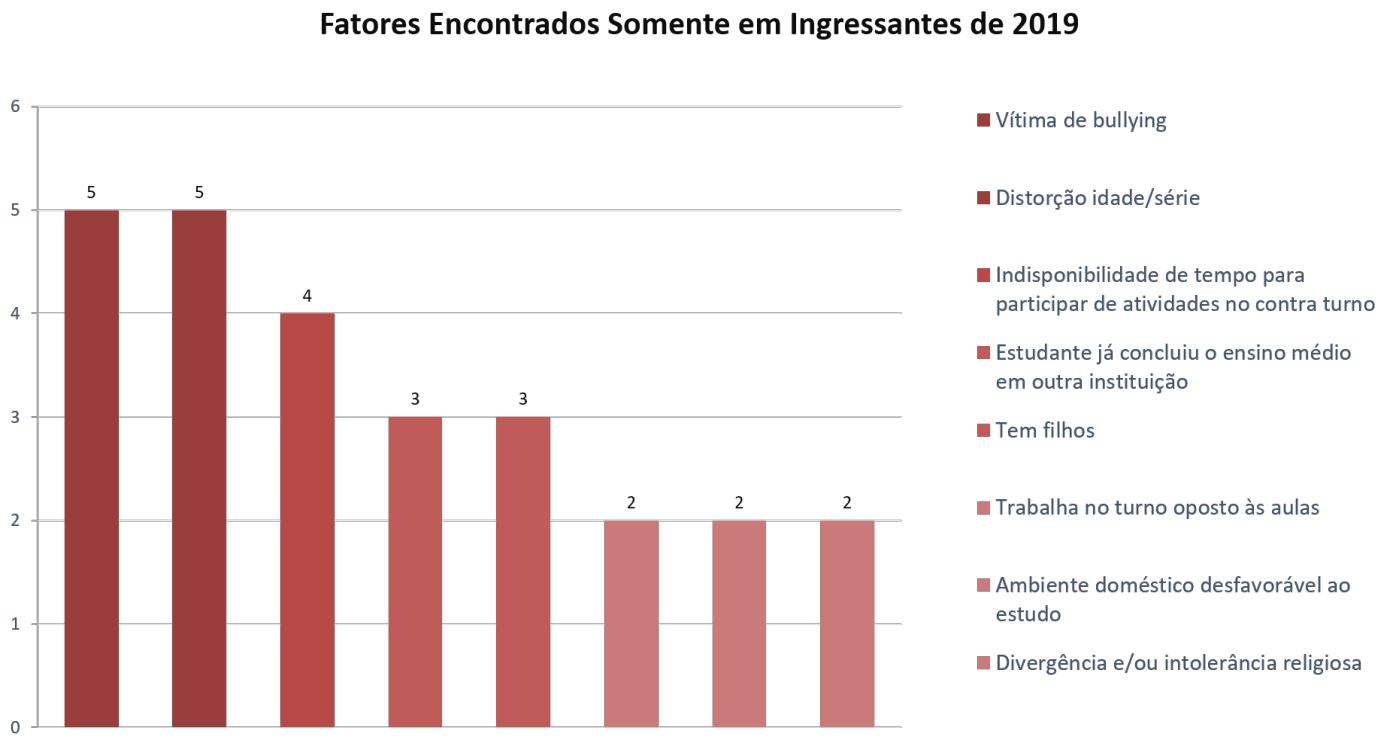

Figura 29: Quantidade de fatores encontrados somente em ingressos de 2019.