

MINERAÇÃO DE DADOS EDUCACIONAIS PARA PREDIÇÃO DA EVASÃO DE ALUNOS NO IFSP CAMPUS ITAPETININGA

MINING EDUCATIONAL DATA TO PREDICT STUDENTS DROPOUT IN A PROFESSIONAL PUBLIC SCHOOL INSTITUTION

EXTRACCIÓN DE DATOS EDUCATIVOS PARA PREDECIR LA DESERCIÓN DE ESTUDIANTES EN UNA INSTITUCIÓN ESCOLAR PÚBLICA PROFESIONAL

Paulo Henrique Vieira Cândido¹

Lucas Tadeu Damaceno²

Carlos Henrique da Silva Santos³

Resumo: A gestão dados escolares vêm recebendo suporte tecnológico e são importantes ferramentas de gestão, tanto da perspectiva educacional quanto administrativa. Educacionalmente, esses dados podem auxiliar no levantamento de problemas de evasão e contribuir na identificação de alunos com perfis para abandono escolar. Assim, este trabalho tem como objetivo levantar e minerar dados para aplicação de aprendizado de máquina no IFSP, com especial atenção ao campus Itapetininga, para se verificar se é possível prever perfis de alunos com possibilidade de evasão escolar. Portanto, justificando-se esse trabalho quanto a tentativa de se maximizar o uso de recursos públicos com redução da evasão escolar e potencializar o êxito dos alunos no campus. Os resultados do trabalho desenvolvido mostraram com certa margem de segurança a predição da evasão escolar com os dados considerados desde o pré-processamento da base de dados do SUAP, perpassando por técnicas de aprendizado de máquina até a visualização dos resultados aqui apresentados. Isso exposto, tem-se um ferramental computacional que pode auxiliar a gestão escolar e, principalmente, fomentar práticas que busquem reduzir o abandono escolar no campus Itapetininga.

Palavras-chave: Evasão Escolar, Aprendizado de Máquina, Mineração de Dados, Dados Escolares, Previsão.

¹ Graduando em Licenciatura em Física. Instituição. ORCID: 0000-0002-2333-9520. E-mail: phvcandido@gmail.com.

² Especialista em Informática Aplicada à Educação. IFSP, campus Itapetininga. ORCID: 0000-0002-3139-5557. E-mail: lucas.bicudo@aluno.ifsp.edu.br

³ Doutor em Engenharia Elétrica. IFSP, campus Itapetininga. ORCID: 0000-0002-8786-405X E-mail: carlos.santos@ifsp.edu.br

Abstract: School data management received technological support and is an important management tool, both from an educational and administrative perspective. Educationally, these data can help to identify students' dropout problems and contribute to the identification of possible profiles for it. Thus, this work aims to collect data and mine for the application of machine learning at IFSP, with special attention to the Itapetininga campus, to verify whether it is possible to predict profiles of students with the possibility of school dropout. Therefore, this work is justified in terms of the attempt to maximize the use of public resources with a reduction in school dropouts and to enhance the success of students on campus. The results of the work carried out showed, with a certain margin of safety, the prediction of school dropout with the data considered from the pre-processing SUAP database, followed by machine learning techniques application and converging to the data visualization results herein presented. Henceforth, it provides a computational tool to support school management and, mainly, encourage practices that seek to reduce school dropout on the Itapetininga campus.

Keywords: School Dropout, Machine Learning, Data Mining, Educational Data, Forecasting.

Resumen: La gestión de datos escolares ha venido recibiendo apoyo tecnológico y es una importante herramienta de gestión, tanto desde el punto de vista educativo como administrativo. Educativamente, estos datos pueden ayudar a identificar problemas de deserción y contribuir a la identificación de estudiantes con perfiles de deserción. Por lo tanto, este trabajo tiene como objetivo recopilar y extraer datos para la aplicación de aprendizaje automático en IFSP, con especial atención al campus de Itapetininga, para verificar si es posible predecir perfiles de estudiantes con posibilidad de deserción escolar. Por lo tanto, este trabajo se justifica en términos del intento de maximizar el uso de los recursos públicos con una reducción de la deserción escolar y mejorar el éxito de los estudiantes en el campus. Los resultados del trabajo realizado mostraron, con cierto margen de seguridad, la predicción de la deserción escolar con los datos considerados del preprocesamiento de la base de datos del SUAP, pasando por técnicas de aprendizaje automática hasta la visualización de los resultados aquí presentados. Según lo expuesto, existe una herramienta computacional que puede ayudar a la gestión escolar y, principalmente, incentivar prácticas que busquen reducir la deserción escolar en el campus de Itapetininga.

Palabras-clave: Abandono escolar, Machine Learning, Minería de datos, Datos escolares, Pronósticos.

Submetido 10/08/2023

Aceito 17/10/2023

Publicado 07/11/2023

Introdução

A atual estrutura educacional brasileira tem diferentes abordagens para o assunto da evasão escolar, perpassando considerações encontradas na literatura desde a educação básica (CRUZ; GONÇALVES, 2015; SILVA FILHO; DE LIMA ARAÚJO, 2017), até as universidades, tanto públicas (FEY; DE CASTILHOS LUCENA; DA SILVA FOGAÇA, 2013; SACCARO; FRANÇA; JACINTO, 2019; SILVA FILHO et al., 2007) quanto as privadas (DA CUNHA et al., 2015; FRITSCH; DA ROCHA; VITELLI, 2015).

Tal preocupação quanto à permanência dos alunos na escola tem suas motivações por parte das universidades. Alguns pensadores que trouxeram grandes colaborações para a educação e que se pode correlacionar a evasão de maneira indireta são Paulo Freire (1921-1997) e Jean Piaget (1896-1980), que reforçaram a importância do interesse do aluno em ir à escola. Assim, pensa-se que se pode associar a evasão escolar e a preocupação com os motivos que corroboram para o seu desenrolar podem seguir alguns pontos desses pensadores.

Freire (FREIRE, 2005) coloca em pauta a importância do diálogo entre o professor e o aluno. Diálogo esse que se torna imprescindível para a vida acadêmica, tornando-se responsabilidade grande do educador em sala de aula. Compreende-se que há de existir interesse do aluno em estar na escola e aprender, para que possa existir um diálogo entre professor e aluno propiciando assim, segundo Freire (FREIRE, 2005), uma experiência mais eficaz que pode levar o aluno para um melhor processo de ensino-aprendizagem.

Piaget (PIAGET, 2013), traz uma perspectiva que pode ter abstrações na mesma linha. Considerando o desinteresse, por parte de Piaget, considera-se a afetividade como aspecto importante para o desenvolvimento da inteligência. Segundo Piaget (PIAGET, 2013), sem o afeto não há o interesse e, por consequência, não há o desenvolvimento, o que propicia a não formação de estruturas cognitivas.

Olhar para essas abordagens, considerando o contexto de desenvolvimento do aluno, perante a sua permanência na escola, leva ao reflexo do que motivou tamanha quantidade de alunos evadidos. Não se coloca em pauta neste trabalho a responsabilidade do professor quanto à sua colocação aberta em sala de aula, compreende-se que a contextualização e formação dos educadores têm praticado isso. Pode-se exemplificar neste ponto a formação de professores e suas práticas educacionais envolvendo tecnologias como ferramenta que busca auxiliar na motivação e envolvimento do aluno do processo de ensino-aprendizagem, assim como práticas

de metodologias ativas de aprendizagem que os tornam os protagonistas de seu aprendizado (SHIGUNOV NETO; MACIEL, 2016).

Neste sentido, uma inquietação acerca de compreender melhor o estado da evasão nas universidades e colaborar no sentido computacional surge. Com tantos recursos tecnológicos e o uso de técnicas de predição de alunos em potencial de evasão, como em (LANES; ALCANTARA, 2018) e (KANTORSKI et al., 2016), apresenta-se a questão de quais resultados se teria na aplicação em institutos da Rede Federal.

A preocupação quanto à evasão no Brasil tem seus esforços iniciados em 1995 por uma comissão especial fomentada e financiada pelo Ministério da Educação (MEC), com o intuito de se analisar o êxito, retenção e abandono dos estudantes em cursos de graduação em instituições públicas (RIGO.; CAZELLA; CAMBRUZZI, 2012). Mais recentemente, (MARQUES; et al, 2019) apresentaram uma revisão de literatura sobre as técnicas e tecnologias que têm sido utilizadas no Brasil para a mineração de dados escolares e com foco na evasão escolar e que em cerca de 75% dos trabalhos utilizados em suas análises a ferramenta Weka (também utilizada neste trabalho) foi a selecionada para a aplicação posterior em aprendizado de máquina para predição dos perfis de alunos em situação de evasão escolar.

Uma das consequências do uso de mineração de dados e o aprendizado de máquina para a evasão escolar vêm sendo o desenvolvimento de ferramentas e formas de visualização desses dados em ontologias e relatórios que buscam facilitar aos gestores escolar seu uso no cotidiano (DA SILVA; RUY; MUTZ, 2022).

Há de se destacar exemplos de estudos que vêm sendo realizados envolvendo dados e o aprendizado de máquina quanto ao impacto que a evasão escolar tem proporcionado durante e após a pandemia (NERI; OSORIO, 2021), sobre a inserção de práticas EaD híbrida ou no ensino totalmente a distância em cursos técnicos (SOARES; et al, 2020) e de curta duração (LIMA, 2019). Também estudos similares realizados na própria Rede Federal como os apresentados por (DE MELO; DE ASSIS MEDEIROS, 2021) quanto a evasão em curso de Licenciatura em Letras no IFPB e no mesmo IF em cursos técnicos subsequentes (DUTRA; DE SOUZA; DE SOUZA FERNANDES, 2022), assim como em dados abertos e disponibilizados pela Plataforma Nilo Peçanha (OLIVEIRA; MEDEIROS; ANDRADE, 2022).

Dessa forma, a fim de corroborar a essa inquietação, este trabalho objetiva-se em oferecer informações que venham fomentar a tomada de decisões no Instituto Federal de

Educação, Ciência e Tecnologia de São Paulo, IFSP Campus Itapetininga, que é a instituição de atuação dos autores.

Evasão Escolar e suas Implicações Educacionais

Na literatura há diferentes definições sobre evasão escolar, sendo necessário adotar uma definição que considere os níveis e âmbitos da educação em que se intenta observá-la, pois há diferenciação entre a Educação Básica e o Ensino Superior (FIGUEIREDO; SALLES, 2017). Além disso, também é necessário considerar a modalidade de ensino (presencial, à distância, modular, por disciplina e outras modalidades) que implicam em significados diferentes para evasão ou abandono escolar (DORE; LÜSCHER, 2011; INEP, 2017).

O Dicionário de Indicadores Educacionais, produzido em 2004 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP, 2004), caracteriza alunos evadidos como aqueles que estiveram matriculados na série *s* no ano *m*, mas que não se encontram matriculados da série *s* ou *s+1* no ano *m+1*. O INEP ainda diferencia os termos “abandono” e “evasão” escolar, apontando que “[...] abandono significa a situação em que o aluno desliga-se da escola, mas retorna no ano seguinte, enquanto na evasão o aluno sai da escola e não volta mais para o sistema escolar” (INEP, 2017).

Essas definições são importantes quando se considera a educação profissional e superior. A Associação Nacional dos Dirigentes de Instituições Federais de Ensino Superior (ANDIFES) caracteriza a evasão como o ato do estudante se desligar do curso superior em situações como: o deixar de matricular-se, desistência oficial do curso e transferência ou reopção de curso. A evasão nessas instituições também é considerada quando o aluno se desliga da instituição e isso é registrado em sistema de forma definitiva ou temporária no ensino superior (ANDIFES; ABRUEM; OTHERS, 1996).

Deste modo, com diferentes abordagens encontradas na literatura, adotou-se como definição de evasão escolar neste trabalho, a construída pela Comissão Especial de Estudos Sobre Evasão nas Universidades Públicas Brasileiras e utilizada por Figueiredo e Salles (FIGUEIREDO; SALLES, 2017), que descreve evasão como sendo a “[...] saída definitiva do aluno de seu curso de origem, sem concluí-lo” (ANDIFES; ABRUEM; OTHERS, 1996).

Por isso, define-se a evasão como sendo a saída inesperada de um aluno de seu curso de origem durante o período regular, mas considerando que a mesma pode acontecer de diferentes formas.

Isso posto, cabe esclarecer o contexto em que a evasão aqui descrita será observada, para isso a próxima seção traz a definição e esclarecimento acerca da Rede Federal de Educação e a geração dos dados utilizados neste trabalho.

Rede Federal de Educação Profissional, Científica e Tecnológica e a Geração de Dados

A Rede Federal de Educação Profissional, Científica e Tecnológica, sendo conhecida como Rede Federal, congregou no ano de 2019 um total de 661 unidades (campus) espalhados por todos os entes federativos do Brasil e divididos em 38 Institutos Federais, 02 Centros Federais de Educação Tecnológica (Cefet), a Universidade Tecnológica Federal do Paraná (UTFPR), 22 escolas técnicas vinculadas às universidades federais e o Colégio Pedro II. A origem dessa importante rede educacional profissional do Brasil se deu em 2008 pela Lei número 11.892⁴.

A estrutura dos Institutos Federais consiste em uma Reitoria e vários campus com gestão interdependente, responsáveis pela execução dos objetivos institucionais. Esses objetivos educacionais surgem a partir da lei de criação dos Institutos Federais, que determina a oferta de pelo menos 50% de vagas para cursos técnicos de nível médio e 20% de formação docente (licenciaturas plenas ou programas especiais de formação docente). O restante das vagas pode ser alocado em outros cursos, como de graduação e pós-graduação, perfazendo em uma única instituição de ensino desde o ensino básico até a pós-graduação. Isso compõe um cenário de gestão complexo tanto da perspectiva legal quanto administrativa para se manter diferentes índices de administração e evasão escolar em cada um desses contextos (DE MARIZ FERNANDES, 2009).

Com a ampliação dessa Rede Federal, tornou-se importante compreender as variações nos índices de evasão escolar, propiciando a implementação de políticas públicas para mitigar esse problema como, por exemplo, o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI) que visa a ampliação do acesso e a permanência dos

⁴ http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111892.htm

estudantes nos Institutos Federais de Educação Superior (IFES) e o Programa Nacional de Assistência Estudantil (PNAES), com contribuição para um melhor desempenho acadêmico e ações preventivas às situações de evasão, como o investimento na alimentação com produtos da região escolar (GILIOLI, 2016).

Assim, dada a complexidade e abrangência dessa rede, a sua gestão fica a cargo da Secretaria de Educação Profissional e Tecnológica do Ministério da Educação (Setec/MEC)⁵. A Setec coleta os dados brutos pelo Sistema Nacional de Informação da Educação Profissional e Tecnológica (SISTEC), para centralizar os dados das instituições e também para a validação de diplomas expedidos pelas instituições da Rede Federal (SETEC, 2018).

Outro sistema que vem sendo implantado pela Setec é a plataforma Nilo Peçanha⁶, que segue o Plano Nacional de Integridade, Transparência e Combate à Corrupção⁷ e a Lei 12.527⁸ que regulamenta o acesso a informações dos governos e autarquias.

Por meio desta plataforma é possível consultar diferentes informações públicas acerca dos estados do Brasil, dentre as quais a quantidade de campus do Instituto Federal. Utilizando deste recurso, constatou-se que no ano de 2018 o Estado de São Paulo dispunha de 32 campus sob a égide do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), sendo 4 avançados e 1 centro de referência. Com base nisso alguns estudos e análises tornam-se possíveis, principalmente para realização de uma comparação entre os diferentes campi do estado de São Paulo.

Assim, visando verificar se os dados extraídos de um campus para comparação com o IFSP como um todo, uma análise desses dados é apresentada na seção de Resultados, em que são dispostos também os resultados de mineração e aplicação de técnicas de aprendizado de máquina para inferência de possibilidade de evasão escolar com os dados coletados até o ano de 2019, que eram os disponíveis à época deste trabalho.

⁵ <http://portal.mec.gov.br/setec-secretaria-de-educacao-profissional-e-tecnologica>

⁶ <http://plataformanilopecanha.mec.gov.br/>

⁷ <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2017/chamada-publica-plano-nacional-de-integridade-transparencia-e-combate-a-corrupcao>

⁸ http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

Mineração de Dados, Aprendizado de Máquina e os Dados Escolares

A Mineração de Dados (Data Mining) é uma área da Ciência da Computação que atua na gestão, processamento e visualização de dados armazenados nos diferentes tipos de Bancos de Dados, trabalhando sobre a complexidade de integração de dados distribuídos em sistemas multiplataforma, tanto com variabilidade quanto volume, que tornam essas atividades ainda mais complexas (CAMILO; SILVA, 2009).

Segundo Navega (2002, p. 1), “[...] apenas recuperar informação não propicia todas as vantagens possíveis. O processo de Data Mining permite que se investigue esses dados à procura de padrões que tenham valor para a empresa”. Deste modo, a Mineração de Dados torna-se um meio que possibilita a identificação de informação útil dentro das organizações, há riqueza de dados e pouca disponibilidade de informações (NAVEGA, 2002).

Assim, o Data Mining torna-se prática importante como ferramenta de suporte estratégico para retenção de clientes, telemarketing, medicina, recursos humanos e dentre outros, por fornecer informação valiosa (CAMILO; SILVA, 2009).

Atentando-se para isso, da perspectiva educacional, este trabalho considera os índices de evasão como importantes fatores administrativos e estratégicos para ações de investimentos, contudo, visando maximizar o sucesso dos alunos em suas trajetórias educacionais.

Isso posto, ao considerar o cenário do IFSP, com sua variedade de cursos e abrangência no Estado de São Paulo, é impraticável tentar se valer de métodos tradicionais de se explorar os dados, em que se depende da análise manual e interpretação por especialistas, buscando por tendências e mudanças nos dados para que sejam elaborados relatórios para uma tomada de decisão futura (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Isso implica na necessidade de migração para modelos mais modernos baseados em Descoberta de Conhecimento em Base de Dados (KDD - Knowledge Discovery in Database) (com suas etapas descritas na Figura 1), para que esse processo seja automatizado perpassando por práticas de Mineração de Dados e Aprendizado de Máquina (COSTA et al., 2019). Nessa Figura 1 percebe-se o fluxo dos dados desde sua coleta num banco de dados, passando pela estruturação e pré-processamento para transformá-los em estruturas compreensíveis as ferramentas computacionais de interesse que os processarão para possibilitar a identificação padrões e fornecimento de visualizações e aplicações baseadas em conhecimentos extraídos dessa variabilidade de dados original.

Figura 1 - Representação do Processo de KDD.



Fonte: O Autor (2021)

Entende-se que a Mineração de Dados é, portanto, a etapa em que os dados são processados para se atribuir significado utilizando técnicas que podem realizar tarefas de Descrição, Classificação, Estimação, Regressão, Agrupamento, Associação e, ou, Predição de dados (FERRARI; SILVA, 2017), sendo esta última o foco desse trabalho.

A predição é a aplicação em que se tem um conjunto de dados já analisados e significado descrito sob certas condições e que seu uso será de prever possíveis cenários ao se ponderar certas variações ainda não analisadas (FERRARI; SILVA, 2017). Quando esses dados estão rotulados são aplicados recursos de aprendizado de máquina supervisionado, usualmente associados à classificação. Em cenários com dados não rotulados é aplicado o aprendizado não supervisionado com técnicas de estimação para se buscar o valor de um certo atributo (CAMILO; SILVA, 2009).

Assim, o aprendizado de máquina é essencialmente a integração entre a estatística e a computação para gerar conhecimento de base de dados que usualmente tem grande volume e variabilidade de dados. Nesta variabilidade inclui-se tanto dados quantitativos quanto qualitativos, que devem ser ressignificados para que sejam inter-relacionados e retornem as respostas de maneira automatizada (BRUCE; BRUCE; GEDECK, 2020).

Esses avanços possibilitaram o surgimento da Mineração de Dados Educacionais, ou *Educational Data Mining* (EDM), que tem sido definida como uma área de pesquisa com o objetivo de desenvolvimento de métodos para explorar conjuntos de dados coletados em ambiente educacional (BAKER; ISOTANI; CARVALHO, 2011). A EDM tem atuado tanto na perspectiva pedagógica, para a customização do ensino quanto na administração da instituição para prover ferramentas que auxiliem na gestão escolar (BAKER; ISOTANI; CARVALHO, 2011; COSTA et al., 2013).

Metodologia e Desenvolvimento

A predição de evasão escolar é uma atividade complexa e requer dados sensíveis dos alunos. Para isso, esse trabalho foi devidamente registrado e aprovado no Comitê de Ética em Pesquisa, sob o protocolo 17438819800005473. Após esse aceite foi possível iniciar as atividades de levantamento de dados educacionais abertos e a mineração de dados disponíveis no Sistema Unificado de Administração Pública (SUAP), que é utilizado pelo IFSP.

Esse levantamento foi tomado como uma análise quantitativa para a exploração de dados disponíveis nesta ferramenta, como uma primeira iniciativa exploratória para se buscar compreender a variabilidade e consistência dos dados disponíveis, buscando correlacioná-los estatisticamente para retornar inferências sobre a possibilidade de evasão escolar de determinados perfis de alunos de acordo com seus dados armazenados no SUAP. Portanto, uma pesquisa quantitativa, exploratória e como um estudo de caso focado na predição da evasão escolar.

Para isso, foi necessário o levantamento de dados educacionais abertos foi realizado previamente à mineração para se ter parâmetros nacionais de comparação, para isso tomou-se como base as informações disponibilizadas na plataforma Nilo Peçanha e que são descritas na próxima seção.

Com esse levantamento, foi feita a coleta de dados da plataforma SUAP, considerando as situações entre os anos de 2014 e 2018 e restringindo-se apenas ao campus Itapetininga, que é a unidade em que os autores atuam.

Essa limitação foi necessária para poder realizar um trabalho de mineração de dados mais efetivo, ao se conhecer a realidade, podendo verificar informações em base de arquivos

não digitalizados até o momento e poder contar com a colaboração de servidores da unidade para verificar se os resultados desse processo estavam corretos.

Além disso, verificou-se ao longo da mineração dos dados que havia erros de inconsistência e diferentes campos em branco nos registros, possivelmente em decorrência do histórico de migração de sistemas ao longo desses 10 anos de atividades na instituição, o que demonstrou mais ainda a necessidade dessa restrição.

Após a etapa de mineração de dados, o aprendizado de máquina foi realizado utilizando a ferramenta Weka, por ser gratuita e disponibilizar diferentes algoritmos (HALL et al., 2009). Além disso, com a mineração dos dados já mencionada, os dados foram organizados em colunas e salvos arquivos Comma-separated values (CSV) com as colunas devidamente rotuladas, para que práticas de aprendizado supervisionado fossem realizadas..

Evasão do IFSP campus Itapetininga

O campus Itapetininga é uma das 37 unidades do IFSP, se localiza na Região Sudoeste do Estado de São Paulo e faz parte da Região Metropolitana de Sorocaba (RMS). Neste campus, até o início de 2014, os cursos oferecidos eram graduação em Licenciatura em Física, técnicos em Edificações, Informática e Mecânica, sendo apenas este último com turmas duplicadas. Para analisar a quantidade de alunos evadidos, considerou-se o número de matriculados entre os anos de 2014 e 2018, por se ponderar que todos os cursos já tinham pelo menos uma turma formada.

Gráfico 1 - Número de matriculados, concluintes, retidos e evadidos entre 2014 e 2018 IFSP, campus Itapetininga.



Fonte: O Autor (2021);

Para iniciar as análises e contextualização nacional da unidade, foram coletados dados da Plataforma Nilo Peçanha, conforme apresentado no Gráfico 1, em que a coluna azul representa o aumento no número de matrículas até o ano de 2017, tendo uma queda no ano de 2018, inclusive no número de concluintes que representam todas as matrículas que tiveram alteração para concluído durante o período anual em laranja.

Nisso, há de se destacar no Gráfico 2 a proporção entre o número de alunos matriculados e concluintes, que é pequeno em todos os anos. Assim, dois critérios podem ser considerados, sendo o primeiro de que matriculados são todos os alunos da unidade e concluintes são aqueles que cumpriram todos os requisitos e se formaram. O segundo ponto é que na licenciatura são 4 anos de curso com entrada anual e os cursos técnicos modulares semestral, semestralmente com entrada de novas turmas.

Por mais que se tenha informações de outros anos para compor a pesquisa, a Plataforma Nilo Peçanha foi implementada em 2017. Assim, a análise dos alunos evadidos se restringe aos anos de 2017 e 2018.

Além disso, serão consideradas na análise as situações de evasão determinadas pela PNP 2018 e 2019, que representam a saída do aluno de seu curso de origem, sem concluí-lo: Abandono, Cancelamento, Desligamento, Transferido Externo, Transferido Interno e Reprovação (INEP, 2017).

No Gráfico 2 tem-se o número de evadidos em cada uma dessas situações de evasão mencionadas, podendo ver a diferença entre 2017 e 2018.

Gráfico 2 - Evadidos por situação de evasão entre 2017 e 2018 IFSP, campus Itapetininga.

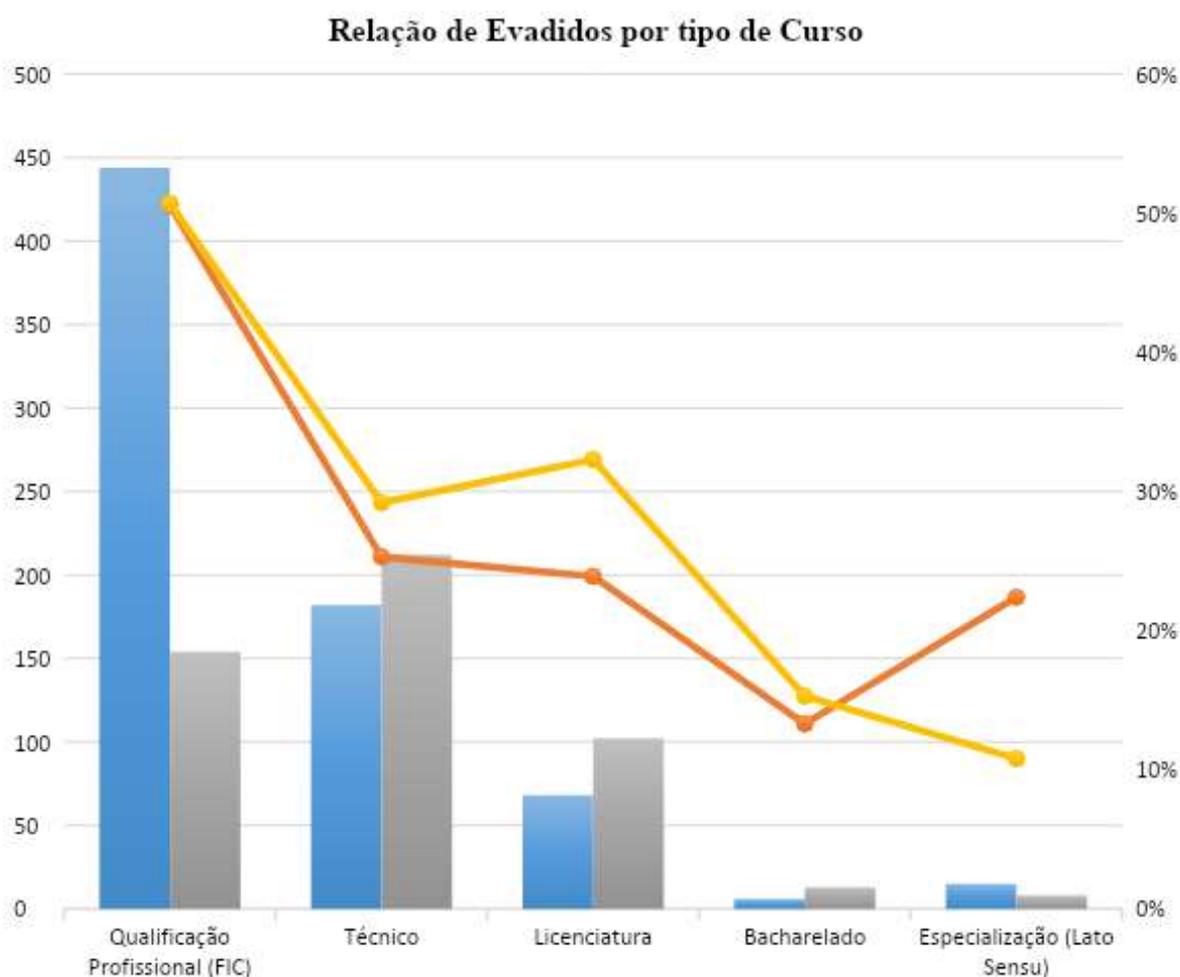


Fonte: O Autor (2021);

Analisando o Gráfico 2, nota-se também que o maior número absoluto de situação de evasão observado é o abandono, que representa os estudantes que não deram nenhum retorno à Coordenadoria de Registros Acadêmicos (CRA) da instituição sobre sua saída.

De tal forma, o Gráfico 3 também corrobora para isso, pois permite observar o número de evadidos e a taxa de evasão por tipo de curso entre 2017 e 2018 através das colunas de cores azul e cinza e a tabela disposta na imagem como resumo de dados.

Gráfico 3 - Relação de evadidos e taxa de evasão por tipo de curso entre 2017 a 2018 do IFSP, campus Itapetininga.

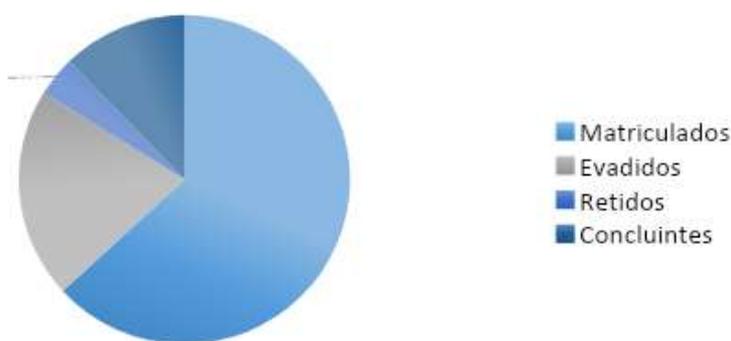


Fonte: O Autor (2021);

Considerando o Gráfico 3, constata-se que os três maiores índices de taxas de evasão foram respectivamente os dos cursos de Qualificação Profissional (FIC), Técnicos e de Licenciatura. Informação que é sintetizada na Gráfico 4, apresentando o percentual de matriculados, retidos, evadidos e concluintes no ano de 2018.

Gráfico 4 - Matriculados, retidos, evadidos e concluintes 2018 IFSP, campus Itapetininga.

Relação entre Matriculados, Evadidos, Retidos e Concluintes



Fonte: O Autor (2021);

Assim, conhecendo os dados acerca do histórico do campus Itapetininga quanto às matrículas, alunos em situação de evasão e outras situações iniciou-se a aplicação dos processos de mineração a fim de desenvolver a predição dessas situações, que é retratada na próxima seção, a de Resultados.

Resultados

Para realização da predição com a realização dos processos de mineração, utilizou-se de uma base de dados construída com os dados disponibilizados pelo SUAP. Houve atenção especial à sensibilidade desses dados e, como supracitado, de acordo com a aprovação do comitê de ética, portanto, não houve nenhum tipo de identificação a fim de manter a integridade de todos os dados dos alunos em questão.

Esse processo foi iniciado utilizando os recursos do SUAP, em que um relatório em planilha para manipulação e filtragem de dados foi exportado. A esse foi aplicada uma filtragem

ao prontuário de todos os alunos matriculados desde o ano de 2010 no campus de Itapetininga. Essa filtragem extraiu 4.453 matrículas até o mês de agosto de 2019, quando os dados foram coletados.

Deste modo, iniciou-se o processo para realização da Mineração dos Dados, sendo a primeira etapa a de seleção, onde buscou-se selecionar atributos relevantes para o estudo da evasão. Nessa seleção, atentou-se aos atributos que se relacionam de alguma forma com as causas internas ou externas que podem levar um aluno ao abandono.

Desta forma, dos 62 atributos utilizados pela base de dados, 15 foram selecionados, sendo estes: Ano de Conclusão do Ensino Anterior, Ano de Ingresso, Cidade, Curso, Data de Matrícula, Data de Nascimento, Estado, Estado Civil, Etnia/Raça, Frequência no Período, Índice de Rendimento do Aluno (IRA), Sexo, Situação no Curso, Tipo de Escola de Origem e Turno.

Após essa primeira etapa de seleção, iniciou-se a etapa de limpeza, em que as redundâncias e valores incompletos dos dados foram retirados, verificando também matrículas repetidas a partir do mesmo prontuário. Foi possível identificar com essa limpeza a relação de 1,53% de valores não preenchidos para Ano de Conclusão Escolar, 1,24% para Cidade, 1,24% para Estado, 8,92% para Estado Civil e 7,28% para Tipo de Escola de Origem.

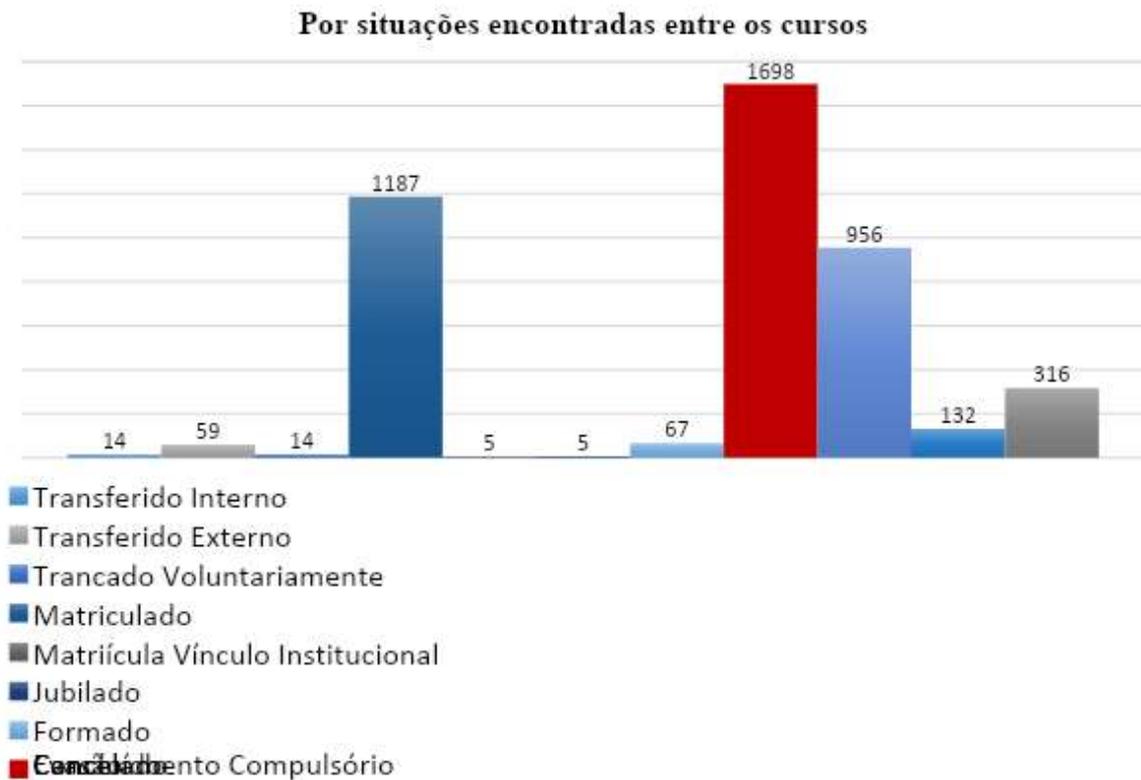
A partir disso, percebeu-se a possível relação e cruzamento de dados para redução de objetos vazios. Redução essa em que foram cruzados o Tipo de Escola de Origem com Instituição de Ensino Anterior e Cidade, Estado e Endereço. Todavia, alguns objetos não são passíveis desse tipo de relação e, para complementar esses dados, foram atribuídos os valores “Não informado” e “Pública” aos atributos Estado Civil e Tipo de Escola de Origem, este último por predominância do valor.

Com esse ajuste dos atributos, tornou-se possível a inserção de novos em uma etapa de transformação, com base nos já existentes, como o Tempo sem Estudar, Idade na Matrícula, Faixa Etária e a Classificação do IRA. Esses atributos finais atribuídos à transformação foram: Cidade, Nome do Curso, Estado, Estado Civil, Etnia, Sexo, Tipo de Escola de Origem, Turno, Faixa Etária, Classificação IRA, Tempo sem estudar, Idade na Matrícula e Situação no Curso. Para conversão dessa base utilizou-se o software Weka.

Após todo esse pré-processamento e a mineração da base de dados extraída do SUAP, foi preciso analisar a informação a fim de extrair conhecimento útil. Neste caso, os processos

da descoberta de conhecimento e a Aprendizagem de Máquina puderam mostrar uma visão geral da situação da evasão no IFSP campus Itapetininga.

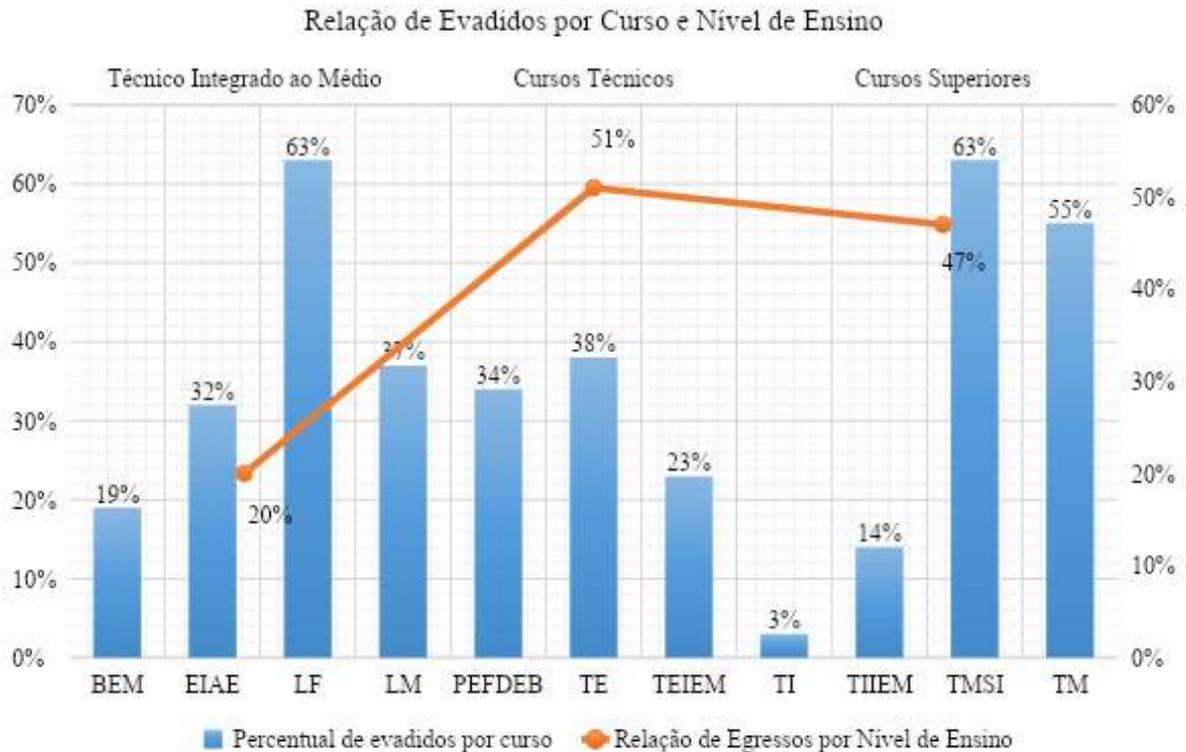
Gráfico 5 - Situação por curso IFSP, campus Itapetininga.



Fonte: O Autor (2021);

Assim, diferentes gráficos e percepções para discussões surgiram. O primeiro destes é o Gráfico 5, que apresenta o número de estudantes em cada situação no curso entre os anos de 2010 a agosto de 2019, esse gráfico foi criado usando todas as 4.453 matrículas registradas.

Gráfico 6 - Percentual de evadidos por curso de 2010 a 2019.



Fonte: O Autor (2021);

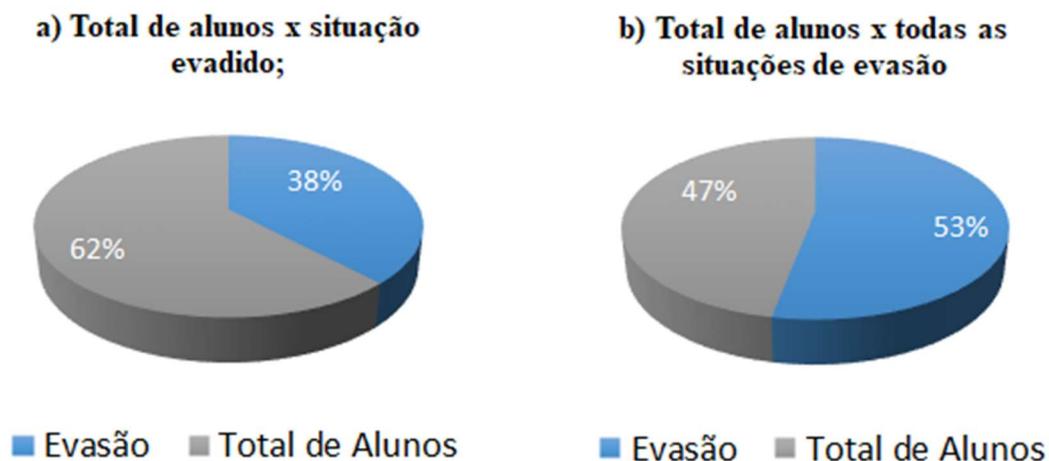
BEM – Bacharelado em Engenharia Mecânica, EIAE – Especialização em Informática Aplicada à Educação, LF – Licenciatura em Física, LM – Licenciatura em Matemática, PEFDEB – Programa Especial de Formação de Docentes para a Educação Básica, TE – Técnico em Edificações, TEM – Técnico em Eletromecânica, TI – Técnico em Informática, TMSI – Técnico em Manutenção e Suporte em Informática, TM – Técnico em Mecânica.

Neste sentido, cabe ressaltar a representação diferente para cada um dos cursos do campus, desde os da educação básica até os de pós-graduação. O Gráfico 6 retrata esse percentual de evadidos em relação aos ingressos de cada curso entre 2010 e 2019.

Ainda, considerando o Gráfico 6, destaca-se que os maiores percentuais de evasão estão nos cursos de Licenciatura em Física (63%), Técnico em Manutenção e Suporte em Informática (63%) e Técnico em Mecânica (55%), enquanto os menores percentuais estão nos cursos de Técnico em Informática (3%), Técnico em Informática Integrado ao Ensino Médio (14%) e Bacharelado em Engenharia Mecânica (19%).

Junto dessas informações do Gráfico 6, também é disposto um gráfico sobreposto em linha, na cor laranja. Esse gráfico representa a porcentagem de evadidos por nível de ensino, evidenciando que nos cursos técnicos há o maior percentual de evasão (51%), enquanto nos cursos integrados ao ensino médio o menor (20%) e os cursos superiores se encontram entremédio destes (47%).

Gráfico 7 - a) Total de alunos x situação evadido;
b) Total de alunos x todas as situações de evasão.



Fonte: O Autor (2021);

A relação desses alunos evadidos é mostrada no Gráfico 7, criando a relação entre o total de alunos, os que estão em situação de evasão, todos os alunos e todas as situações de evasão. Dos 4.453 alunos, 1.698 são evadidos, o que representa 38% do total $\frac{1.698}{4.453} \cdot 100 = 38$.

Todavia, considerando outras situações de abandono de um curso, tais como a situação cancelado (316), os transferidos internos (14) e externos (59) e os alunos que trancaram a matrícula voluntariamente (14), este número sobe para 2.101 (1.698 + 316 + 14 + 59 + 14 = 2.101) o que representa 47% do total de alunos ($\frac{2.101}{4.453} \cdot 100 = 47$), como visto nos gráficos na Gráfico 7.

Neste sentido, também cabe considerar que esses dados aconteceram em decorrência de diferentes anos. Assim, observar a evasão através dos anos se torna interessante. Por conta

disso, apresenta-se no Gráfico 8 o percentual de evadidos anualmente de ingresso entre 2010 e 2018 considerando todas as situações de evasão (2.101).

Gráfico 8 - Percentual de evadidos por ano de ingresso.



Fonte: O Autor (2021);

Ao analisar o Gráfico 8 pode-se perceber que o número de evadidos diminuiu entre 2010 e 2013, tendo um pico no ano de 2014 e voltando a reduzir até o ano de 2018 com 28%. Nota-se também que em determinados anos o percentual se manteve o mesmo (2011 e 2012; 2013 e 2016) e nos anos de 2017 e 2018 ele se manteve abaixo da média (49%). Tendo observado a relação por curso, por ano e os diferentes níveis de ensino, cabe atentar-se às características dos alunos dispostos nos registros. Por conseguinte, o Gráfico 9 mostra o percentual de evadidos por sexo, considerando o total de ingressos do sexo masculino em relação ao total de evadidos, o mesmo aplicado para o sexo feminino entre os anos de 2010 a agosto de 2019.

Gráfico 9 - Percentual de evadidos por sexo.



Fonte: O Autor (2021).

Outro conhecimento sensível e de interesse extraído dos dados dos dados minerados é a evasão por raça ou etnia. Para isso, é interessante também se verificar a frequência relativa ocorrida em cada uma delas e quando totalizada no percentual da população amostral como indicado na Tabela 1. O status "Não Declarado" associa-se a pessoas que não informaram quando foram se cadastrar e também aos registros que estavam em branco no banco de dados, talvez por isso a variação mais acentuada nesse tipo em que a quantidade preenchida com as regras aqui indicadas tenham contribuído para esse percentual bastante acima da média das demais raças/etnias. Outro dado interessante de observar é de que no geral as raças/etnias ficaram em torno de 41% de evasão, tendo um salto apenas nos cadastrados como Amarelo que percentualmente ficaram com 57,32% de evasão.

Tabela 1. Valores absolutos e relativos à Evasão Escolar por Raça/Etnia

Raça/Etnia	Ingressos	f(%)	Evasão	f(%)	% de Evasão por Raça/Etnia
Amarela	82	1,84%	47	2,24%	57,32%
Branca	2628	59,02%	1114	52,02%	42,39%
Indígena	10	0,22%	4	0,19%	40,00%
Parda	121	2,72%	53	2,52%	43,80%
Preta	121	2,72%	53	2,52%	43,80%
Não Declarada	917	20,59%	598	28,46%	65,21%

O município de Itapetininga apresenta uma extensão territorial bastante grande, sendo o terceiro maior município do Estado de São, segundo dados apurados no IBGE, estando atrás de Iguape e Itapeva apenas. Ponderando sua extensão e analisando a evasão escolar, os percentuais tanto em moradores do município quanto os externos a eles foram de 47%, não havendo variação observável nesse contexto. Quanto ao estado civil, havia apenas cerca de 10% dos alunos com informações cadastradas e nesse subconjunto amostral extraiu-se que 7% e 9% dos casados e solteiros evadiram, respectivamente. Portanto, não sendo possível inferir se há alguma relação de acordo com o valor desses atributos.

Outro dado interessante de considerar é a evasão quanto a alunos originários de escolas públicas e privadas. Dos mais de 4000 estudantes cadastrados, apenas 2005 tinham essa informação cadastrada, sendo que desses 43% dos alunos de escolas privadas evadiram contra 46% dos alunos de escolas públicas, sendo uma variação relativamente baixa e que também acaba por não ser um fator implicante na análise de evasão com os dados coletados.

Nesse cenário, observa-se que a faixa etária seria um prognóstico de cuidado à evasão. Para se determinar intervalos considerou-se crianças até 14 anos, de 15 a 24 anos como jovem, dos 25 aos 64 anos de idade como adulto e dos 65 anos em diante como idoso. Da população total observada tem-se que houve a inscrição de 191 crianças e que delas 46 evadiram (24%), quanto aos jovens foram 2851 ingressantes com 1302 evasões (46%), 1410 inscrições de adultos com 753 evasões (53%), e apenas um idoso que se inscreveu e concluiu seu curso, portanto sem evasão nesta última categoria. Com todo esse levantamento apresentado é possível estabelecer um perfil de predominância estatística de aluno evadido por curso do campus no período analisado. Esses perfis são apresentados na Tabela 2.

Tabela 2. Perfis de Evasão Escolar por curso e associando sexo, etnia/raça, estado civil e faixa etária

Curso	Sexo	Etnia/Raça	Estado Civil	Faixa Etária
Esp. Info. Apl. Edu	M	Parda	Casado	Adulto
Bach. Eng. Mecânica	F	Parda	Casado	Jovem

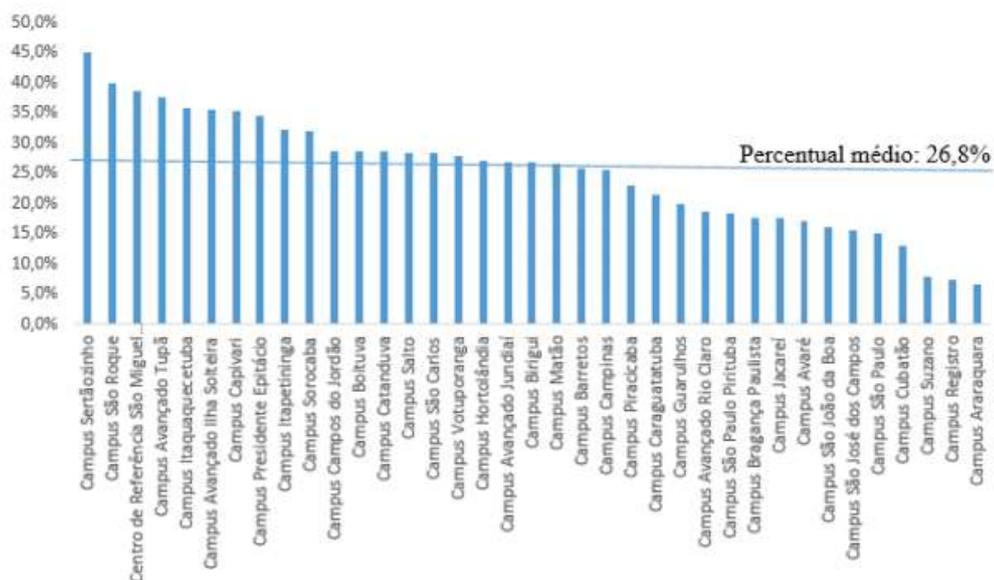
Lic. Física	M	Amarela	Solteiro	Adulto
Lic. Matemática	F	Preta	Solteiro	Adulto
Prog. Form. Docente	M	Amarela	Solteiro	Jovem
Téc. Edificações	M	Amarela	Indiferente	Adulto
Téc. Int. Eletromecânica	M	Preta	Solteiro	Criança
Téc. Informática	F	Branca	Solteiro	Jovem
Téc. Int. Informática	M	Branca	Solteiro	Jovem
Téc. M.S.I.	F	Amarela	Indiferente	Adulto
Téc. Mecânica	F	Amarela	Solteiro	Adulto
Campus Itapetininga no Geral	M	Amarela	Solteiro	Adulto

Esses perfis foram extraídos da população de dados analisados e considerando os que mais se repetiam em cada curso, de acordo com a disponibilidade de dados. Deve-se considerar nessa observação da Tabela 2 que para cada curso há diferentes quantitativos absolutos de alunos inscritos e respectivamente evadidos, sendo que desses os que mais se repetem são indicados nesta tabela. Para se estimar de maneira simplista o perfil do aluno que mais evade no campus, independente de curso na última linha dessa tabela foi considerado apenas o tipo que mais repetia nas linhas anteriores para os diferentes cursos então em valores percentuais e absolutos da base de dados para se diminuir a quantidade de ponderações que deveriam ser realizadas nos dados. Assim, percebe-se que o perfil de aluno do sexo Masculino da Etnia/Raça Amarela, solteiro e na faixa etária adulta é que tende a ter maior chance de evasão no campus.

Considerando o foco deste trabalho em analisar e inferir evasão no campus Itapetininga como estudo de caso e validação de metodologia adotada, pensando-se na possibilidade de estender depois essas análises aos demais 36 campus do IFSP, um comparativo de dados abertos de evasão extraídos da Plataforma Nilo Peçanha demonstra que a média evasão do campus

Itapetininga se está acima da maioria das demais unidades, como indicado no Gráfico 10. Não como se realizar detalhamento estatístico como o apresentado aqui com as demais unidades neste trabalho, pois não era o foco e a legalidade para sua ocorrência, tanto da perspectiva ética científica quanto ao atendimento da LGPD estão em curso para que seja posteriormente realizada para todo o IFSP, respeitando privacidade e anonimização dos dados a serem coletados.

Gráfico 10 - Taxa de evasão por campus do IFSP.



Ainda, atentando-se aos níveis demonstrados no Gráfico 10, percebe-se a maior taxa de evasão no campus Sertãozinho, próximo de 45,0%, enquanto a menor está no campus Araraquara. Também interessante ressaltar que o percentual médio de taxa de evasão entre todos os campus é de 26,8%, o campus Itapetininga ocupa o 9º lugar com 32,3% de evasão, isso evidencia que está acima da média de evasão e é um dos 10 campus com maior taxa de evasão em todo o estado.

Conclusões

Este trabalho objetivou-se buscar informações que pudessem ser úteis para tomada de decisões no IFSP Campus Itapetininga quanto à evasão dos estudantes. Para isso, foram exploradas as formas de evasão, suas diferentes definições; as formas de aprendizado de máquina e a mineração de dados, suas etapas e como esta poderia ser aplicada com dados educacionais; por fim, com as informações encontradas através das análises e o processo de revisão construído, foi possível criar dados sobre a evasão escolar e suas características no IFSP Campus Itapetininga, permitindo análises acerca do total, situação por curso, nível de ensino e sexo entre as evasões de 2010 até 2019.

Através do processo construído é possível criar hipóteses e certezas sobre características dos cursos e suas taxas de evasão, um exemplo disso é a alta taxa de evasão do curso Técnico em Manutenção e Suporte em Informática, que pode ser observada no Gráfico 6.

Utilizando essas etapas foi possível a realização do que fora objetivado, conseguindo informações concisas, precisas e desenvolvidas com base em sistematização, com possível aplicação de predição dos alunos evadidos do IFSP Campus Itapetininga. Portanto, considera-se que decisões de ações da instituição podem ser realizadas com base nas informações dispostas na seção de resultados.

Com isso, alguns trabalhos futuros podem ser notados, como a disponibilização de uma página da Web ou aplicativo que forneça esses dados em tempo real, considerando principalmente que isso demandaria de consumo dos dados do SUAP e plataforma Nilo Peçanha. A integração dessas informações ao SUAP e a divulgação desse método a ser adotado em outros campi do Instituto Federal ou outras universidades também. Outro trabalho futuro a este compreende a análise da trajetória desses alunos quanto às experiências pessoais que culminam na evasão escolar.

Referências

ANDIFES, ANDIFES; ABRUEM, ABRUEM; OTHERS. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas: resumo do relatório apresentado a ANDIFES, ABRUEM e SESu/MEC pela Comissão Especial. Avaliação: **Revista da Avaliação da Educação Superior**, [S. l.], v. 1, n. 2, 1996. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/225423/000224712.pdf?sequence=1>. Acesso em 20 de agosto de 2022.

BAKER, Ryan; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de dados educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, [S. l.], v. 19, n. 02, p. 3, 2011. Disponível em: <http://ojs.sector3.com.br/index.php/rbie/article/view/1301>. Acesso em 20 de agosto de 2022.

BRASIL, Ministério da Educação. Secretaria de Educação Profissional e Tecnológica. Brasília, 2018 Disponível em: <http://portal.mec.gov.br/rede-federal-inicial/instituicoes>. Acesso em 20 de agosto de 2022.

BRUCE, Peter; BRUCE, Andrew; GEDECK, Peter. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. [s.l.] : **O'Reilly Media**, 2020.

CAMILO, Cássio Oliveira; SILVA, João Carlos Da. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFC)**, [S. l.], p. 1–29, 2009. Disponível em: https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF_001-09.pdf. Acesso em 20 de agosto de 2022.

COSTA, Claudio Napolis; COUTINHO, Jonatas Vieira; DE MAGALHÃES, Lúcia Helena; ARBEX, Márcio Aarestrup. Descoberta de conhecimento em bases de dados. **Revista Eletrônica: Faculdade Santos Dumont**, [S. l.], v. 2, 2019. Disponível em: <https://www.fsd.edu.br/wp-content/uploads/2019/12/artigo9.pdf>. Acesso em 20 de agosto de 2022.

COSTA, Evandro; BAKER, Ryan SJD; AMORIM, Lucas; MAGALHÃES, Jonathas; MARINHO, Tarsis. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, [S. l.], v. 1, n. 1, p. 1–29, 2013. Disponível em: <http://ojs.sector3.com.br/index.php/pie/article/view/2341>. Acesso em 20 de agosto de 2022.

CRUZ, Erica; GONÇALVES, Márcia Ribeiro. Evasão na educação de jovens e adultos. **LINKSCIENCEPLACE - Interdisciplinary Scientific Journal**, [S. l.], v. 2, n. 3, 2015.

DA CUNHA, Jacqueline Veneroso Alves; DE LUCA, Márcia Martins Mendes; DE LIMA, Gerlando Augusto Sampaio Franco; CORNACCHIONE JR, Edgard Bruno; OTT, Ernani. Quem está ficando para trás? Uma década de evasão nos cursos brasileiros de graduação em Administração de Empresas e Ciências Contábeis. **Revista de Educação e Pesquisa em Contabilidade**, [S. l.], v. 9, n. 2, p. 124–142, 2015. Disponível em: <https://www.redalyc.org/pdf/4416/441642788002.pdf>. Acesso em 20 de agosto de 2022.

DA SILVA, Eduardo Moura; RUY, Fabiano Borges; MUTZ, Filipe Wall. Abordagem para análise de múltiplas fontes de dados de evasão escolar. **Anais do Computer on the Beach**, v. 13, p. 149-156, 2022. Disponível em: <https://periodicos.univali.br/index.php/acotb/article/view/18791/10792> Acessado em 03 de julho de 2023.

DE MARIZ FERNANDES, Francisco das Chagas. Gestão dos Institutos Federais: O desafio do centenário da rede federal de educação profissional e tecnológica. **HOLOS**, [S. l.], v. 2, p. 3–9, 2009. Disponível em: <https://www.redalyc.org/pdf/4815/481549226002.pdf>. Acesso em 20 de agosto de 2022.

DE MELO, João Ricardo Freire; DE ASSIS MEDEIROS, Ana Karoliny. Evasão escolar precoce no ensino superior a distância: Uma análise segundo os dados do curso de licenciatura em letras no IFPB. **Research, Society and Development**, v. 10, n. 7, p. e31510717230-e31510717230, 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/download/17230/15016>. Acesso em 03 de julho de 2023.

DORE, Rosemary; LÜSCHER, Ana Zuleima. Permanência e evasão na educação técnica de nível médio em Minas Gerais. **Cadernos de pesquisa**, [S. l.], v. 41, n. 144, p. 770–789, 2011. Disponível em: <https://www.scielo.br/j/cp/a/jgRKBkHs5GrxxwkNdNNtTfM/abstract/?lang=pt>. Acesso em 20 de agosto de 2022.

DUTRA, Janderson Ferreira; DE SOUZA, João Paulo Lopes; DE SOUZA FERNANDES, Damires Yluska. Classificação de estudantes com potencial à evasão: aplicando mineração de dados no contexto de cursos técnicos subsequentes do IFPB. **Revista Principia-Divulgação Científica e Tecnológica do IFPB**, v. 59, n. 3, p. 1009-1027, 2022. Disponível em: <https://periodicos.ifpb.edu.br/index.php/principia/article/download/5488/1795>. Acesso em 03 de julho de 2023.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, [S. l.], v. 17, n. 3, p. 37, 1996. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>. Acesso em 20 de agosto de 2022.

FERRARI, DANIEL GOMES; SILVA, LEANDRO NUNES D. E. CASTRO. **Introdução a mineração de dados**. São Paulo, Saraiva, 2017.

FEY, Ademar Felipe; DE CASTILHOS LUCENA, Karina; DA SILVA FOGAÇA, Valéria Nagali. Evasão no Ensino Superior: uma pesquisa numa IES do ensino privado. **Revista de Humanidades, Tecnologia e Cultura**, [S. l.], v. 1, n. 1, 2013. Disponível em: <http://www.fatecbauru.edu.br/rehutec/artigos/3-Evas+o%FAo%20no%20Ensino%20Superior.pdf>. Acesso em 20 de agosto de 2022.

FIGUEIREDO, Natália Gomes da Silva; SALLES, Denise Medeiros Ribeiro. Educação Profissional e evasão escolar em contexto: motivos e reflexões. **Ensaio: Avaliação e Políticas Públicas em Educação**, [S. l.], v. 25, n. 95, p. 356–392, 2017. Disponível em: <https://www.scielo.br/j/ensaio/a/Bw8WKpzdP3w8qn5zL68C3sq/abstract/?lang=pt>. Acesso em 20 de agosto de 2022.

FREIRE, Paulo. **Pedagogia da autonomia: saberes necessários à prática educativa**. São Paulo: Paz e Terra, 1996.

FRITSCH, Rosângela; DA ROCHA, Cleonice Silveira; VITELLI, Ricardo Ferreira. A evasão nos cursos de graduação em uma instituição de ensino superior privada. **Revista Educação em Questão**, [S. l.], v. 52, n. 38, p. 81–108, 2015. Disponível em: <https://periodicos.ufrn.br/educacaoemquestao/article/view/7963/>. Acesso em 20 de agosto de 2022.

GILIOLI, Renato de Sousa Porto. Evasão em instituições federais de ensino superior no Brasil: expansão da rede, Sisu e desafios. **Brasília: Câmara dos Deputados**, [S. l.], p. 49, 2016. Disponível em: https://nupe.blumenau.ufsc.br/files/2017/05/evasao_institui%C3%A7%C3%B5es.pdf. Acesso em 20 de agosto de 2022.

HALL, Mark; FRANK, Eibe; HOLMES, Geoffrey; PFAHRINGER, Bernhard; REUTEMANN, Peter; WITTEN, Ian H. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, [S. l.], v. 11, n. 1, p. 10–18, 2009. Disponível em: <https://dl.acm.org/doi/abs/10.1145/1656274.1656278>. Acesso em 20 de agosto de 2022.

INEP, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Dicionário de indicadores educacionais: fórmulas de cálculo**. Coordenação-Geral de sistemas integrados de informações educacionais, Brasília 2017.

KANTORSKI, Gustavo; FLORES, Evandro G.; SCHMITT, Jáder; HOFFMANN, Ivan; BARBOSA, Fernando. Predição da evasão em cursos de graduação em instituições públicas. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE) 2016, **Anais** [...]. [s.l: s.n.] p. 906. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/6776>. Acesso em 20 de agosto de 2022.

LANES, Mariele; ALCÂNTARA, Cleber. Predição de Alunos com Risco de Evasão: estudo de caso usando mineração de dados. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE) 2018, **Anais** [...]. [s.l: s.n.] p. 1921. Disponível em: <http://ojs.sector3.com.br/index.php/sbie/article/view/8191>. Acesso em 20 de agosto de 2022.

LIMA, Monique Tamara de. **Aplicação de técnicas de aprendizado de máquina para classificar alunos de cursos de idiomas com relação à possibilidade de evasão**. 2019. Trabalho de Conclusão de Curso. Universidade Tecnológica Federal do Paraná. Disponível em: http://riut.utfpr.edu.br/jspui/bitstream/1/12285/1/LD_COENP_2019_1_09.pdf. Acessado em: 03 de julho de 2023.

MARQUES, Leonardo Torres et al. Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. **RENOTE**, v. 17, n. 3, p. 194-203, 2019. Disponível em: https://www.researchgate.net/profile/Caetano-Segundo/publication/320994972_Minerao_de_dados_aplicada_a_educacao_um_estudo_comparativo_acerca_das_caracteristicas_que_influenciam_a_evasao_escolar/links/5c4ef51aa6fdccd6b5cf333a/Minerao-de-dados-aplicada-a-educacao-um-estudo-comparativo-acerca-das-caracteristicas-que-influenciam-a-evasao-escolar.pdf. Acesso em 02 de julho de 2023.

NERI, Marcelo; OSORIO, Manuel Camillo. Evasão escolar e jornada remota na pandemia. **Revista NECAT-Revista do Núcleo de Estudos de Economia Catarinense**, v. 10, n. 19, p. 28-55, 2021. Disponível em: <https://www.academia.edu/download/89214841/479044992.pdf>. Acessado em: 03 de julho de 2023.

SHIGUNOV NETO, Alexandre; MACIEL, Lizete Shizue Bomura. Refletindo sobre o passado, o presente e as propostas futuras na formação de professores. **Revista Internacional de Formação de Professores**, v. 1, n. 2, p. 172-186, 2016.

NAVEGA, Sergio. Princípios essenciais do data mining. **Anais do Infoimagem**, [S. l.], 2002. Disponível em: <http://www.intelliwise.com/reports/i2002.pdf>. Acesso em 20 de agosto de 2022.

OLIVEIRA, Isleimar S.; MEDEIROS, Francisco Petrônio A.; ANDRADE, Fabio G. Seleção de Atributos para Classificadores de Evasão Escolar com Dados da Plataforma Nilo Peçanha. In: **Anais do I Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil**. SBC, 2022. p. 30-39. Disponível em: <https://sol.sbc.org.br/index.php/wapla/article/download/22520/22344/> Acessado em: 03 de julho de 2023.

PIAGET, Jean. **A psicologia da inteligência**. Petrópolis. Vozes, 2013.

RIGO, Sandro J.; CAZELLA, Silvio C.; CAMBRUZZI, Wagner. Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In: **Anais do Workshop de Desafios da Computação Aplicada à Educação**, 2012. p. 168-177. Disponível em <https://www.academia.edu/download/54682137/2787-4763-1-SM.pdf> . Acessado em 03 de julho de 2023.

SACCARO, Alice; FRANÇA, Marco Túlio Aniceto; JACINTO, Paulo de Andrade. Fatores Associados à Evasão no Ensino Superior Brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de Ciência, Matemática e Computação e de Engenharia, Produção e Construção em instituições públicas e privadas. **Estudos Econômicos (São Paulo)**, [S. l.], v. 49, n. 2, p. 337–373, 2019. Disponível em: <https://www.scielo.br/j/ee/a/9YxHxWkk6Dzy35CpgmxXbPt/abstract/?lang=pt>. Acesso em 20 de agosto de 2022.

SILVA FILHO, Raimundo Barbosa; DE LIMA ARAÚJO, Ronaldo Marcos. Evasão e abandono escolar na educação básica no Brasil: fatores, causas e possíveis consequências. **Educação por escrito**, [S. l.], v. 8, n. 1, p. 35–48, 2017. Disponível em: <https://revistaseletronicas.pucrs.br/index.php/porescrito/article/view/24527>. Acesso em 20 de agosto de 2022.

SILVA FILHO, Roberto Leal Lobo; MOTEJUNAS, Paulo Roberto; HIPÓLITO, Oscar; LOBO, Maria Beatriz de Carvalho Melo. A evasão no ensino superior brasileiro. **Cadernos de pesquisa**, [S. l.], v. 37, n. 132, p. 641–659, 2007. Disponível em: <https://www.scielo.br/j/cp/a/x44X6CZfd7hqF5vFNnHhVWg/abstract/?lang=pt>. Acesso em 20 de agosto de 2022.

SOARES, Leandra Cristina Cavina Piovesan et al. Aplicação de técnicas de aprendizado de máquina em um contexto acadêmico com foco na identificação dos alunos evadidos e não evadidos. **Humanidades & Inovação**, v. 7, n. 8, p. 223-235, 2020. Disponível em: <https://revista.unitins.br/index.php/humanidadesinovacao/article/download/3293/1617> Acessado em 03 de julho de 2023.