

REGRA DE ASSOCIAÇÃO COM O APRIORI PARA DETECTAR O PERFIL DO ALUNO EVASOR NO CURSO DE LICENCIATURA EM COMPUTAÇÃO

ASSOCIATION RULES WITH APRIORI TO DETECT THE EVASOR STUDENT PROFILE IN COMPUTER LICENSE COURSE

Gabriel Rosa **1**
Diego Rodrigues **2**

Resumo: Este artigo visa abordar a problemática da evasão aplicando técnicas de Mineração de Dados em um conjunto de informações coletadas a respeito de todos os acadêmicos já matriculados no curso de graduação em Licenciatura em Computação ofertado pelo Instituto Federal de Educação, Ciências e Tecnologia do Tocantins (IFTO) – Campus Dianópolis. Mineração de dados ou Data Mining é o processo de explorar grandes quantidades de dados utilizando algoritmos específicos na busca de padrões consistentes e conhecimentos válidos. O principal objetivo desta obra está atrelado em utilizar a linguagem R (R Studio) e seu algoritmo de regras de associação para identificar o perfil do aluno evasor matriculado entre o período de 2015 a 2018.

Palavras-chave: Evasão. Licenciatura em Computação. Mineração de Dados.

Abstract: This article aims to address the problem of dropout by applying Data Mining techniques to a set of information collected about all students already enrolled in the undergraduate degree in Computing offered by the Federal Institute of Education, Science, and Technology of Tocantins (IFTO) - Dianópolis Campus. Data Mining is the process of exploring large amounts of data using specific algorithms in search of consistent standards and valid knowledge. The main objective of this work is to use the R (R Studio) language and its association rules algorithm to identify the student evasion profile enrolled from 2015 to 2018.

Keywords: Computer Degree. Data Mining. School dropout.

Licenciado em Computação, IFTO. Lattes: <http://lattes.cnpq.br/1352568169280215>. ORCID: <https://orcid.org/0000-0003-2600-2832>.
E-mail: gabriel.lcc.rosa@gmail.com

Mestre em Modelagem Computacional de Sistemas, UFT. Lattes: <http://lattes.cnpq.br/9582612733423890>. ORCID: <https://orcid.org/0000-0001-8396-1947>. E-mail: diego.rodrigues@ifto.edu.br

Introdução

A evasão é um problema que assola a educação nacional em todos os níveis, desde a educação básica até os cursos superiores. Este fenômeno ocorre quando os alunos perdem o vínculo com a instituição antes do período de conclusão. Pode ocorrer por diversos motivos como dificuldades relacionadas ao processo de ensino-aprendizagem, problemas financeiros, trabalho, saúde, não identificação com o curso, transferência de instituição e inúmeros outros casos.

Essa realidade não está distante do Instituto Federal de Educação Ciências e Tecnologia do Tocantins (IFTO) – Campus Dianópolis que possui um elevado índice de evasão no curso de Licenciatura em Computação desde a implantação do curso até os dias atuais.

Este trabalho tem a proposta de utilizar a Ciência de Dados por meio das técnicas de Mineração de Dados através do algoritmo de regra de associação apriori, para definir padrões e tendências dos alunos evadidos do IFTO – Campus Dianópolis baseado em dados históricos dos acadêmicos matriculados no curso durante o período de 2015 a 2018 no curso de Licenciatura em Computação.

Mineração de dados é um processo de tratamento de grandes quantidades de dados para encontrar padrões e correlações, ou seja, consiste de técnicas para extração de dados. Mineração de dados ou Data Mining é o termo que passou a ser adotado em 1990. Em sua composição contém disciplinas que caminham entrelaçadas: Estatística (coleta e análise de dados que obtém resultados numéricos), inteligência artificial e machine learning (aprendizado de máquina).

O principal desafio desta pesquisa é expor como as ferramentas de Mineração de dados podem ser úteis para as Instituições de Ensino Superior (IES) para municiar a equipe gestora na detecção de padrões e tendências, e como essa técnica pode ser eficaz no combate à evasão.

Neste relato de experiência, será descrito todo o percurso desta monografia que teve como metodologia analisar dados documentais fornecidos pela Coordenação de Registros Escolares (CORES) aplicando técnicas de Mineração de Dados a fim obter uma melhor interpretação dos resultados utilizando abordagens estatísticas.

Evasão nos cursos de licenciatura

A licenciatura é uma das modalidades de ensino superior que permite ao estudante graduado exercer o magistério na área de conhecimento em que se formou, nos seguimentos de ensino da educação básica, ou seja, ensino fundamental, ensino médio e ensino profissionalizante.

De acordo com o decreto instituído na RESOLUÇÃO CNE/CP 1, DE 18 DE FEVEREIRO DE 2002, os cursos na modalidade licenciatura:

Institui Diretrizes Curriculares Nacionais para a Formação de Professores da Educação Básica, em nível superior, curso de licenciatura, de graduação plena.

Art. 1º As Diretrizes Curriculares Nacionais para a Formação de Professores da Educação Básica, em nível superior, em curso de licenciatura, de graduação plena, constituem-se de um conjunto de princípios, fundamentos e procedimentos a serem observados na organização institucional e curricular de cada estabelecimento de ensino e aplicam-se a todas as etapas e modalidades da educação básica.

Art. 3º A formação de professores que atuarão nas diferentes etapas e modalidades da educação básica observará princípios norteadores desse preparo para o exercício profissional específico, que considerem:

I - a competência como concepção nuclear na orientação do curso;

II - a coerência entre a formação oferecida e a prática esperada do futuro professor, tendo em vista:

a) a simetria invertida, onde o preparo do professor, por ocorrer em lugar similar àquele em que vai atuar, demanda consistência entre o que faz na formação e o que dele se espera;

b) a aprendizagem como processo de construção de conhecimentos, habilidades e valores em interação com a realidade e com os demais indivíduos, no qual são colocados em uso capacidades pessoais;

c) os conteúdos, como meio e suporte para a constituição das competências;

d) a avaliação como parte integrante do processo de formação, que possibilita o diagnóstico de lacunas e a aferição dos resultados alcançados, consideradas as competências a serem constituídas e a identificação das mudanças de percurso eventualmente necessárias.

III - a pesquisa, com foco no processo de ensino e de aprendizagem, uma vez que ensinar requer, tanto dispor de conhecimentos e mobilizá-los para a ação, como compreender o processo de construção do conhecimento. (RESOLUÇÃO CNE/CP 1, DE 18 DE FEVEREIRO DE 2002)

Sobre as outras modalidades de ensino superior (bacharelado e Tecnólogo) a RESOLUÇÃO CNE/CP 1, DE 18 DE FEVEREIRO DE 2002, instaura que:

1. Ambos os cursos são em nível superior de graduação plena;
2. Ambos oferecem conhecimento e habilidades distintas;
3. Ambos ensinam intervenções profissionais distintas e específicas, por conseguintes habilitações diferenciadas e impeditivas uma em relação à outra.

Cada curso de licenciatura tem sua própria identidade curricular, entretanto, independente da área de conhecimento, o que toda licenciatura tem em comum é a extensa carga horária referente às disciplinas de cunho pedagógico, voltadas para as técnicas de ensino, psicologia da educação, avaliação no processo de ensino e aprendizagem e didática.

Antes de se matricular em um curso superior é fundamental ter traçado um plano de carreira, pois a principal razão de escolher um curso de licenciatura é lecionar (atuar como docente), ou seja, optar por trabalhar na área pedagógica. No entanto, se não for de interesse atuar no sistema de ensino ou como docente existe outras modalidades de ensino que serão mais indicadas para atender as expectativas de cada indivíduo como o bacharelado.

A grande maioria das pessoas encara a reflexão sobre suas carreiras apenas como identificação de oportunidades e a busca de seu aproveitamento. Ao proceder desta maneira, subordina suas carreiras a uma realidade dada pelo ambiente e perde a condição de atuar sobre a realidade. A atuação sobre o ambiente, no sentido de transformá-lo para melhor adequar-se às preferências e características pessoais, exige que as pessoas, antes de refletirem sobre as oportunidades de carreira oferecidas pelo ambiente, olhem para dentro de si,

procurando identificar o que gostam, o que as satisfaz, o que as estimula, etc. Ao fazê-lo, por estarem usando um padrão interior de avaliação, terão melhor condição para identificar oportunidades de carreira. (DUTRA, 1996, p. 24)

Diante do exposto, as modalidades de graduação: Licenciatura e Bacharelado terão um enfoque maior nesta pesquisa.

Steinhilber (2006), em seu artigo (Licenciatura e/ou Bacharelado: Opções de graduação para intervenção profissional) constata que:

“LICENCIATURA: a formação de professores que atuarão nas diferentes etapas e modalidades da educação básica, portanto, para atuação específica e especializada.

BACHARELADO: qualificado para analisar criticamente a realidade social, para nela intervir por meio das diferentes manifestações.” (STEINHILBER, 2006, p. 19 – 20)

Agora que foi esclarecido o quê a licenciatura representa no cenário da educação superior nacional, pode-se ter um vislumbre da procura de cursos para essa modalidade.

Tabela 1 – Matrículas por Modalidade

Total de Matrículas:		7.037.688
Matrícula por Modalidade	Bacharelado	4.703.804
	Licenciatura	1.366.588
	Tecnólogo	944.764
	Não Aplicável	22.532
Modalidade de Ensino	Presencial	5.923.832
	A Distância	1.113.850

Fonte: INEP (2012)
Elaboração do autor.

Ao observar o número de alunos matriculados nos cursos de graduação no ano de 2011 por meio do censo da educação superior (2012) disponibilizado pelo INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP), é possível notar que a modalidade que concentra a maior quantidade de alunos é o bacharelado, tendo 3 vezes mais alunos matriculados que na modalidade de licenciatura.

Tabela 2 – Estatística das Matrículas por Modalidade

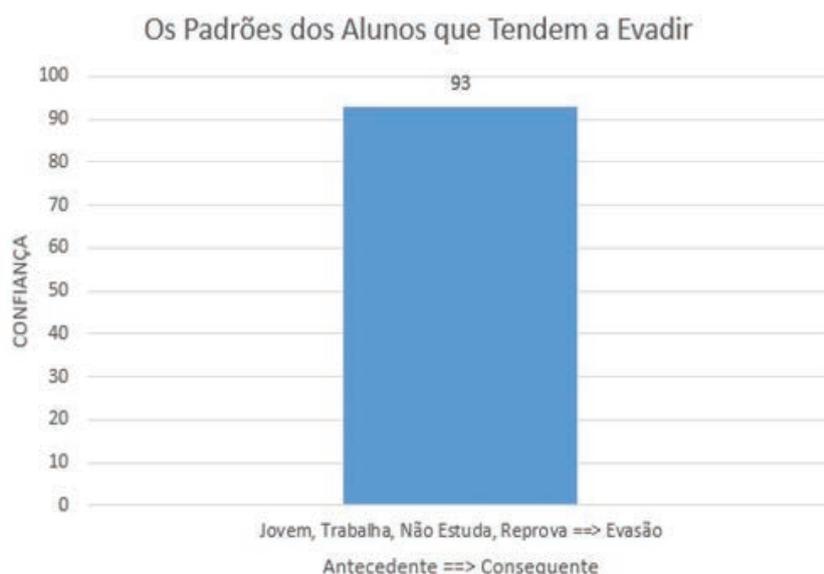
Matrícula por Modalidade	
Bacharelado	66,84%
Licenciatura	19,42%
Tecnólogo	13,42%
Não Aplicável	0,32%

Fonte: INEP (2012)
Elaboração do autor.

Ao analisar a tabela 2, fica ainda mais evidente a depreciação que ocorre nos cursos de licenciatura, pois a nível nacional as vagas ocupadas por essa modalidade não atingem nem 20% da população acadêmica. E como já dito, ingressar em um curso superior nem sempre garante o êxito educacional, ou seja, ainda existe o “fantasma” da evasão que assombra os cursos superiores.

Durante uma pesquisa executada no IFTO – campus Dianópolis no ano de 2017 foi feito um levantamento de dados por meio de um estudo de caso para identificar qual era o perfil dos discentes propensos a evadirem da instituição.

Figura 1 – IFTO padrão de evasão encontrado



Fonte: (ROSA, SOBRINO & RODRIGUES; 2018).

Ao analisar a figura 1, é possível notar que o padrão dos alunos propensos a evadirem eram os estudantes que trabalhavam, não tinham tempo para estudar fora da instituição e já reprovaram em alguma matéria.

Especificamente analisando a situação dos cursos de licenciatura existem outros aspectos que elavam os índices de evasão, como: A desvalorização do Docente, A baixa remuneração e O desprestígio social da profissão. (VITELLI, 2014, p.4)

Mineração de dados

Mineração de dados ou Data Mining é o processo de explorar grandes quantidades de dados utilizando algoritmos específicos na busca de padrões consistentes. Em sua composição contém disciplinas que caminham entrelaçadas: Estatística (coleta e análise de dados que obtém resultados numéricos), Inteligência Artificial e Machine Learning (algoritmos que permitem o computador aprender).

Frequentemente, mineração de dados tem sido considerada e classificada como uma mistura de pesquisas em estatística, inteligência artificial e bancos de dados. Até recentemente, não era reconhecido como um campo de interesse para os estatísticos, sendo mesmo considerado, nesta área, como uma área de pesquisa 'pouco relevante'. Devido à sua importância prática, entretanto, o campo tem emergido como uma área de crescimento acentuado e de elevada importância, destacando-se pelo surgimento de diversos congressos científicos e produtos comerciais. (CORTÉS, PORCARO & LIFSCHITZ, 2002, p. 2)

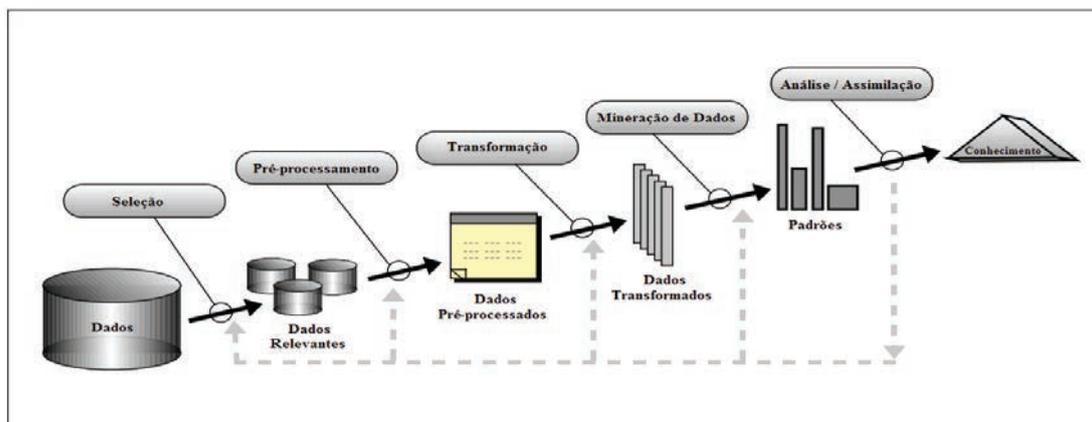
Em relatório publicado pela Gartner Mineração de dados é considerada como “O processo de descobrir correlações, padrões e tendências significativos analisando grandes quantidades de dados armazenados em repositórios. A mineração de dados emprega tecnologias de reconhecimento de padrões, bem como técnicas estatísticas e matemáticas.”

Atualmente, com a grande quantidade de dados gerados e armazenados diariamente é muito difícil e em alguns casos chega a ser impossível analisar manualmente um grande conjunto de dados armazenados em banco de dados. Por este motivo, a mineração de dados tem se tornado uma ferramenta imprescindível, pois em sua essência existe o processo de Descoberta de Conhecimento em Banco de Dados ou Knowledge Discovery in Databases (KDD).

“O volume e a velocidade de dados gerados vêm atingindo níveis sem precedentes. Para se ter uma ideia da dimensão disso, 90% dos dados disponíveis no mundo foram gerados nos últimos dois anos”, afirma Rachel Horta, CEO da Hekima, empresa que faz análise de Big Data. Na escola, a coleta de dados é feita por professores e gestores em situações como a atribuição de notas, a realização dos cálculos de frequência, o preenchimento do Censo e a aplicação de avaliações padronizadas. “Em um primeiro momento, reunir esses indicadores servia para prestar contas à sociedade e mostrar índices de desempenho. A ideia se encerrava aí. Entretanto, essas informações são preciosas e começam a ser entendidas como base para uma avaliação diagnóstica do ensino e para a tomada de decisões”, comenta Luis Antonio Laurelli, diretor executivo da Tuneduc, empresa de Big Data voltada ao universo educacional. (GRINBERGAS, 2015)

Para Fayyad et al. (1996) a busca da Descoberta do Conhecimento em Banco de Dados é um processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados.

Figura 2 – Etapas da descoberta do conhecimento



Fonte: Educ. rev. vol.32 no.1 Belo Horizonte Jan./Mar. 2016.

Para Attar (1999), entre as técnicas de mineração de dados as regras de associação ocupam o segundo lugar em percentual de utilização e aplicações, pois auxiliam no suporte a decisão em sistemas de diagnósticos, além de detectar padrões e tendências eminentes.

Análise de associação é o processo de interconexão de objetos na tentativa de expor características e tendências. Gera redes de interações e conexões presentes nos conjuntos de dados usando as associações item a item. Entende-se que a presença de um item implica necessariamente na presença do outro na mesma transação. O banco de dados é visto como uma coleção de transações, cada uma envolvendo um conjunto de itens. Essas regras correlacionam a presença de um conjunto de itens com um outro intervalo de valores para um outro conjunto de variáveis. (CORTÊS, PORCARO & LIFSCHITZ, 2002, p. 16 – 17)

A tarefa de associação tem o intuito de identificar associações entre registros de dados que, de alguma maneira, estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação. (VASCONCELOS & CARVALHO, 2004, p.07)

A gama de possibilidades que o conjunto de mecanismos e técnicas ofertadas pelas ferramentas de mineração de dados podem proporcionar aos usuários são inúmeras, no entanto é necessário compreender o seu conceito e o tipo de algoritmo que melhor se enquadra a sua base de dados.

Linguagem R

R é uma Linguagem de Programação que disponibiliza diversos recursos estatísticos e de visualização gráfica, a ferramenta foi desenvolvida na década de 90 sendo um software livre de ambiente de desenvolvimento, ou seja, é gratuito para os usuários.

R é uma linguagem e ambiente para computação estatística e gráficos. É um projeto GNU similar à linguagem e ambiente S que foi desenvolvido na Bell Laboratories (anteriormente AT & T, agora Lucent Technologies) por John Chambers e colegas. (R-PROJECT)

Existem diversas Interfaces Gráficas do Utilizador (GUI) para R sendo as mais populares: JGR, Rkward, SciViews-R, Rcmdr e R Studio. O que populariza a ferramenta e faz com que seja muito utilizada por cientistas de dados e estatísticos.

Muitos usuários pensam em R como um sistema de estatísticas. Preferimos pensar nisso como um ambiente no qual as técnicas estatísticas são implementadas. R pode ser estendido (facilmente) através de pacotes. Existem cerca de oito pacotes fornecidos com a distribuição R e muitos outros estão disponíveis na família CRAN de sites da Internet, abrangendo uma ampla gama de estatísticas modernas. (R-PROJECT)

Regras de associação – algoritmo apriori

Apriori é um algoritmo que faz parte do conjunto de algoritmos que compõe as Regras de Associação em Mineração de Dados. Segundo **Arbex, Saboredo e Miranda** “Ele foi proposto pela equipe de pesquisa QUEST da IBM que deu origem ao Software Intelligent Miner. Este algoritmo faz recursivas buscas no Banco de Dados à procura dos conjuntos frequentes”.

Para Baranauskas esse algoritmo pode ser dividido em duas etapas:

- Encontrar os itemsets frequentes (aqueles com suporte $\geq \text{min_sup}$)
- Gerar as regras a partir dos itemsets frequentes (com confiança $\geq \text{min_conf}$)

Para o funcionamento do algoritmo apriori é necessário configurar alguns parâmetros fundamentais para sua execução que são: suporte e confiança.

- Suporte: pode ser definido como a probabilidade de que uma transação qualquer satisfaça o antecedente e o consequente;
- Confiança: pode ser interpretada como a probabilidade de encontrar o consequente da regra nas transações sobre a condição que essas transações também contenham antecedente.

Itemset é composto por todos os itens presentes e relacionáveis que se cruzam em alguma transação. Estes Itemset são ordenados de maneira lexicográfica (ordem alfabética), para relacionar a recorrência dos itens em uma estrutura matemática seguindo a ordem natural de dois conjuntos ordenados no plano cartesiano.

Sua formula pode ser definida em três passos:

Passo 1: Correlação

- $X \Rightarrow Y$ (significa que se X for verdadeiro então Y também é verdadeiro quando $X \subset I$, $Y \subset I$, $X \cap Y = \emptyset$.)
- $X \subset I$ (quer dizer que todos os elementos do conjunto X estão contidos no conjunto I , ou seja, X é um subconjunto de I .)
- $Y \subset I$ (esta formula possui o mesmo significado da anterior.)
- $X \cap Y = \emptyset$ (esta formula representa a intercessão do conjunto X com Y que não pode retornar um conjunto **Nulo**, ou seja, representa um conjunto de elementos que aparecem simultaneamente e pertencem a dois ou mais conjuntos distintos.)

Passo 2: Confiança

- $X \Rightarrow Y$ (só é válido no conjunto de transações (T), se o nível de confiança (C) corresponder em ($C\%$) das transações de $T \supset X$ e $T \supset Y$.)
- $T \supset X$ (significa que o conjunto T contém os elementos do conjunto X .)
- $T \supset Y$ (possui o mesmo significado descrito acima.)

Passo 3: Suporte

- $X \Rightarrow Y$ (o suporte é medido por (S) em T , quando $(S\%)$ de $T \supset XUY$.)
- $T \supset XUY$ (quer dizer que, o conjunto T contém os elementos da união de X e Y .)

Metodologia

O presente material é fruto de uma pesquisa estatística de cunho exploratória, que se fundamenta na Observação (estudo em que os elementos não são afetados) de dados relacionados as características dos alunos do curso de Licenciatura em Computação no IFTO – Campus Dianópolis, abordando três técnicas estatísticas fundamentais: Descritiva, Probabilidade e Inferência.

É descritiva pois os dados coletados foram organizados, resumidos e demonstrados para uma melhor compreensão. Probabilística, pois define os padrões e tendências dos alunos evasores, além de prever quantos alunos que ingressaram em 2019/1 podem evadir. E Inferência, pois por meio desta pesquisa pode-se obter respostas sobre o fenômeno da evasão com dados representativos a Instituição que serão demonstrados nos **Resultados**.

As variáveis utilizadas na pesquisa assumem duas formas, sendo a primeira Qualitativa Nominal, pois os dados coletados na instituição são categóricos e não possuem hierarquia. Já a segunda forma assumida pela variável é Quantitativa Contínua, pelos resultados retornados possuírem valores numéricos reais que podem assumir qualquer intervalo.

A população (alvo do estudo) da pesquisa se deu por todos os alunos que já foram matriculados no curso de Licenciatura em Computação no IFTO – Campus Dianópolis desde a implementação do curso até o período de 2019/1. A amostra (subconjunto da população) estudada são os alunos que ingressaram entre 2015 a 2018, para poder definir os padrões e perfis dos alunos evadidos da instituição.

Para a elaboração deste artigo, o primeiro método utilizado foi a pesquisa bibliográfica em livros, artigos e outros tipos de publicações científicas, além de sites oficiais e especializados, para construir o embasamento teórico desta pesquisa visando fomentar a perspectiva de vários autores a respeito da evasão no curso de Licenciatura em Computação e das técnicas de Mineração de Dados.

A pesquisa documental foi realizada por meio de um conjunto de dados fornecido pela CORES em formato Comma-separated values (CSV) contendo dados dos acadêmicos que ingressaram no IFTO – Campus Dianópolis desde o período de 2015 até 2019. Os dados fornecidos são sigilosos e não possuem nenhuma característica que possa identificar algum aluno, ou seja, de maneira que nenhuma personalidade seja publicada como por exemplo: telefone, nome, número de documentos.

Os dados solicitados e de interesse da pesquisa são:

- Raça;
- Sexo;
- Renda;
- Ano de Nascimento (Faixa etária: Jovem, Adulto, Idoso);
- Ano de Ingresso;
- Percentual de Conclusão;
- Status (Evadido, Em Curso, Concluído)

Para o Ano de nascimento do aluno, foi utilizada uma abordagem diferente, os dados foram classificados e renomeados para Faixa Etária em que foram separadas em três grupos: jovem, adulto e idoso.

- Jovem: Baseado na Lei nº 12.852, sancionada em 5 de agosto de 2013:

Art. 1º Esta Lei institui o Estatuto da Juventude e dispõe sobre os direitos dos jovens, os princípios e diretrizes das políticas públicas de juventude e o Sistema Nacional de Juventude - SINAJUVE.

§ 1º Para os efeitos desta Lei, são consideradas jovens as

pessoas com idade entre 15 (quinze) e 29 (vinte e nove) anos de idade.

- **Idoso:** De acordo com a Lei nº 10.741, sancionada em 1 de outubro de 2003, em seu Art. 1.º É instituído o Estatuto do Idoso, destinado a regular os direitos assegurados às pessoas com idade igual ou superior a 60 (sessenta) anos.

- **Adulto:** Com base na definição da faixa etária dos jovens ser dos 15 aos 29 anos, e dos idosos ser a partir dos 60 anos, os adultos foram classificados com a faixa etária entre os 30 e 59 anos de idade.

Os dados coletados foram utilizados com o intuito de conhecer e compreender os aspectos históricos dos alunos matriculados no curso de Licenciatura em Computação desde a implementação até 2018, para um amplo conhecimento sobre os padrões e tendências dos acadêmicos evasores da Instituição.

Para o processo quantitativo contínuo, foi utilizado a linguagem R por meio da GUI R Studio para desenvolver o algoritmo de regra de associação visando encontrar padrões, tendências e correlações válidas e concisas. No ambiente de desenvolvimento R Studio foi executado o algoritmo apriori que consiste em descobrir padrões associativos, além de possuir capacidade de executar um grande número de atributos, extraindo como resultado inúmeras alternativas combinatórias.

Para a execução do apriori no R Studio foi preciso carregar duas bibliotecas sendo a primeira o pacote “arules” para gerar as regras de associação, e o segundo pacote “arulesViz” para gerar gráficos ou imagens representativas como a classificação lexicográfica exposta na seção de **Resultados**.

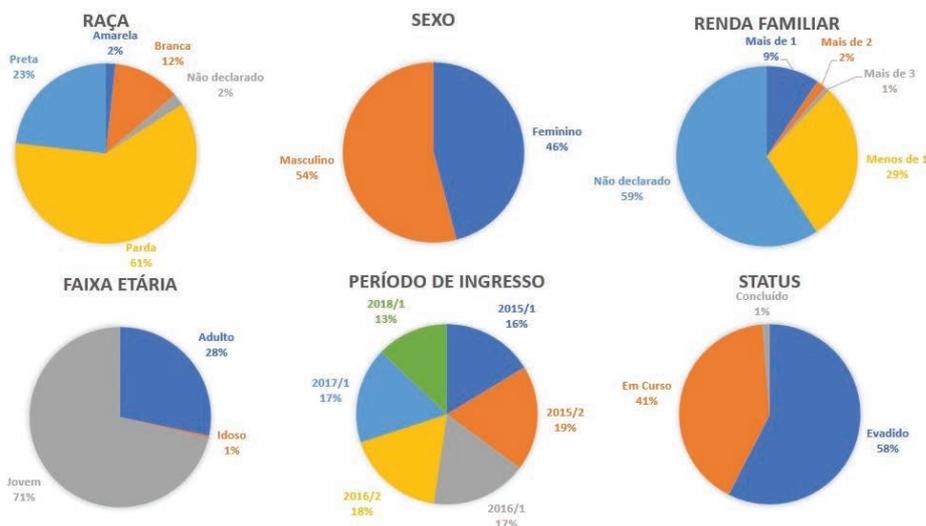
Os parâmetros escolhidos para melhor execução do algoritmo apriori foram os seguintes valores: suporte (supp=0.03) e a confiança (conf=0.8). Foi utilizado outro parâmetro o (minlen=6) para definir o mínimo de características a serem relacionadas a fim de reduzir o volume de dados e informações, expondo apenas correlações que possuíssem o conjunto completo de atributos: sexo, raça, renda, faixa etária, percentual de conclusão e status.

Após a etapa de configuração dos parâmetros do algoritmo apriori foi o momento de execução, onde retornou 10 regras que foram exportadas para um arquivo de texto e armazenadas em planilhas do Excel para serem analisadas.

Resultados

Ao analisar os dados coletados dos acadêmicos ingressantes no IFTO – Campus Dianópolis entre 2015 a 2018 fica evidente a situação de vulnerabilidade socioeconômica de muitos alunos.

Figura 3 – Resumo dos Dados Coletados na Instituição (2015 a 2018).



Fonte: Elaboração do Autor.

O conjunto de respostas da figura 3 representa a amostra de 233 alunos que ingressaram entre 2015 e 2018, o que expõe a situação socioeconômica de muitos acadêmicos, onde, aproximadamente 30% declararam no ato da matrícula que a renda familiar bruta per capita é de menos de um salário mínimo. Existe outro fator que pode acrescentar consideravelmente esse percentual, pois 59% dos acadêmicos não declararam renda.

Ao executar o algoritmo apriori, o mesmo retorna à classificação lexicográfica que mostra em forma de imagem a relação das transações executadas pelo algoritmo conforme pode ser visualizada abaixo:

Figura 4 – Classificação Lexicográfica.



Fonte: Elaboração do Autor.

Para realizar as análises de dados coletados, foram levados em considerações alguns aspectos sobressalentes. Então, optou-se por agrupar as respostas que continham o consequente definido como “Evadido”. Da amostra, portanto, foram geradas 10 transações, mas foram selecionadas apenas 5 regras de associação que atendiam aos requisitos de escolha para serem analisadas e revelar os padrões e tendência dos alunos do IFTO – Campus Dianópolis que evadem da instituição.

Das 5 regras selecionadas foi utilizado o método estatístico de medida de altura em conjunto de dados (Moda), para observar o índice de recorrência e então poder definir o padrão e perfil do aluno evasor.

Dessa forma, os dados nos quais foram concentrados os esforços para realizar as análises desta monografia expressam cinco características relacionais que se destacam e evidenciam os padrões e perfis dos alunos evadidos da instituição.

Tabela 3 – Perfis dos Acadêmicos Evasores.

Antecedente						Consequente	
Sexo	Raça	Renda	Faixa Etária	% conclusão		Status	Confiança
Feminino	Pardo	Não declarada	Adulto	Menos de 20%	⇒	Evadido	100%
Feminino	Pardo	Menos de 1 S/M	Jovem	Menos de 20%	⇒	Evadido	100%
Feminino	Pardo	Não declarada	Jovem	Menos de 20%	⇒	Evadido	100%
Masculino	Pardo	Não declarada	Adulto	Menos de 20%	⇒	Evadido	100%
Masculino	Pardo	Não declarada	Jovem	Menos de 20%	⇒	Evadido	100%

Fonte: Elaboração do Autor.

Pode-se observar por meio da tabela 3 que, que existem cinco padrões para os alunos evadidos, sendo verídico pela atestação da confiança que em todos os casos está em 100%, e pela amostra que evidencia cinco perfis distintos sem recorrência nos padrões da evasão, o que torna este conjunto amodal.

Considerações Finais

Após analisar criticamente todos os dados e seções desenvolvidas desta pesquisa, pôde-se constatar que a Mineração de Dados neste trabalho se mostrou uma ferramenta imprescindível, pois, por meio dela foi possível fazer um diagnóstico da situação do curso de Licenciatura em Computação analisando pontualmente o contexto histórico do curso.

Através do algoritmo de regra de associação foi possível analisar todo o contexto histórico do curso desde a implantação até 2018. Isso se faz necessário pois desde os primórdios a história está presente na sociedade para nos ajudar a esclarecer fatos ocorridos através do tempo. E neste caso não foi diferente, pois, por meio dos dados coletados dos acadêmicos foi possível definir o padrão dos alunos da instituição que tendem a evadir da instituição.

A escolha desse algoritmo foi estratégica, pois com o conhecimento histórico adquirido por meio deste é possível auxiliar a gestão educacional na tomada de decisão para traçar estratégias preventivas a fim de minimizar e reduzir os elevados índices de evasão.

Referências

ARBEX, Eduardo Compasso; SABOREDO, Alexandre de Paiva; MIRANDA, Dhalila. **Implementação e Estudo de caso do algoritmo Apriori para Mineração de Dados**. Disponível em: <https://www.aedb.br/seget/arquivos/artigos04/193_artigo_SEGET.doc>

ATTAR, Akeel Al. **Data mining – Beyond Algorithms**. 1998.

BARANAUSKAS, José Augusto. **Regras de Associação**. Disponível em: <<http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Regras-Associacao.pdf>>

BRASIL. **LEI N.º 10.741, DE 1.º DE OUTUBRO DE 2003**. Disponível em: <http://bvsmms.saude.gov.br/bvs/publicacoes/estatuto_idoso_2ed.pdf>

_____. **LEI Nº 12.852, DE 5 DE AGOSTO DE 2013**. Disponível em: <http://prattein.com.br/home/images/stories/Juventude/Estatuto_da_Juventude.pdf>

_____. **RESOLUÇÃO CNE/CP Nº 1, DE 18 DE FEVEREIRO DE 2002**. Disponível em: <http://portal.mec.gov.br/seesp/arquivos/pdf/res1_2.pdf>

CORTÊS, Sérgio da Costa; PROCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de Dados – Funcionalidades, Técnicas e Abordagens**. PUC-RioInf.MCC10/02 Maio, 2002.

DUTRA, J. S. **Administração de carreiras: uma proposta para repensar a gestão de pessoas**. São Paulo: Atlas, 1996.

FAYYAD, U. M., Piatetsky Shapiro, G., Smyth, P. & Uthurusamy, R. **Advances in Knowledge Discovery and Data Mining**. AAAIPress, The Mit Press, 1996.

Gartner report. **Data Mining**. Disponível em: <<https://www.gartner.com/it-glossary/data-mining>>.

GRINBERGAS, Daniella. Como o Big Data pode ser usado na educação: Como o uso de técnicas de análise de dados, escolas podem mapear o desempenho de seus alunos e melhorar a gestão administrativa, controlando problemas como a evasão. **Revista Educação**, edição 223, 2015.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Censo da educação superior 2011. Brasília: **Inep**; MEC, 2011. Disponível em: <<http://goo.gl/tYbnLd>>.

ROSA, Gabriel; SOBRINHO, Isaac P. F.; RODRIGUES, Diego. **O Uso de Ferramentas de Mineração**

de dados como auxílio na Prevenção da Evasão nas Universidades. 8ª JICE, 2017. Disponível em: <<http://propi.ifto.edu.br/ocs/index.php/jice/8jice/paper/view/8190>>

R-PROJECT. **O Projeto R para Computação Estatística.** Disponível em: <<https://www.r-project.org/>>

STEINHILBER, Jorge. **Licenciatura e/ou Bacharelado: Opções de graduação para intervenção profissional.** Disponível em: <http://www.fae.br/2009/arquivos/educacaofisica_licenciatura-oubacharelado.pdf>.

VASCONCELOS, Livia Maria Rocha de; CARVALHO, Cedric Luiz de. **Aplicação de Regras de Associação para Mineração de Dados na Web.** Technical Report, 2004. Disponível em: <http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf>

Recebido em 28 de setembro de 2019.

Aceito em 30 de março de 2020.