

**MINERAÇÃO DE DADOS NO COMBATE À EVASÃO ESCOLAR EM  
INSTITUIÇÕES DE ENSINO SUPERIOR BRASILEIRAS**

***DATA MINING IN THE FIGHT AGAINST SCHOOL EVASION IN BRAZILIAN  
HIGHER EDUCATION INSTITUTIONS***

Ronaldo Rodrigues Martins – ronaldo.martins@fatec.sp.gov.br  
Faculdade de Catanduva (Fatec) – Catanduva – SP – Brasil

Matheus Henrique Bertuci – matheus.bertuci@fatec.sp.gov.br  
Faculdade de Catanduva (Fatec) – Catanduva – SP – Brasil

Lucas Polimeno Peniani - lucas.peniani@fatec.sp.gov.br  
Faculdade de Catanduva (Fatec) – Catanduva – SP – Brasil

**DOI: 10.31510/inf.v17i2.885**

Data de publicação: 18/12/2020

**RESUMO**

Instituições brasileiras de ensino superior, públicas ou privadas, tem buscado nos últimos anos meios de frear o aumento da evasão escolar. Neste sentido, a mineração de dados tem se mostrado uma ferramenta eficiente para extração de conhecimento de grandes bases de dados, permitindo dessa maneira coordenar ações administrativas e pedagógicas no combate à evasão escolar. Este trabalho apresenta uma revisão de trabalhos primários sobre o uso de mineração de dados no combate à evasão escolar no ensino superior brasileiro. São destacadas as técnicas de mineração, variáveis mais relevantes, acurácia, metodologia, e quais bases de dados são utilizadas na extração de conhecimento. Ao final, são apresentadas direções para trabalhos futuros que podem orientar novas pesquisas sobre o tema.

**Palavras-chave:** Data Mining, Evasão Escolar, Ciência de Dados

**ABSTRACT**

Brazilian institutions of higher education, public or private, have been looking for ways to curb the increase in school dropout in recent years. In this sense, data mining has proved to be an efficient tool for extracting knowledge from large databases, thus allowing to coordinate administrative and pedagogical actions to combat school dropout. This paper presents a review of primary studies on the use of data mining to combat school dropout in Brazilian higher education. Mining techniques are highlighted, most relevant variables, accuracy, methodology, and which databases are used in the extraction of knowledge. At the end, directions for future work are presented that can guide further research on the topic.

**Keywords:** Data Mining, School Dropout, Data Science

## 1. INTRODUÇÃO

De acordo com Lobo (2017), a evasão em instituições de ensino superior (IES), tanto em instituições privadas como em instituições públicas, vem sendo uma questão chave, porém, sem grandes mudanças nos últimos anos, sendo que a taxa permanece na média de 20%. A alta taxa de evasão ocorre em muitos casos devido algumas IESs direcionarem grande parte das suas energias na captação de alunos, deixando em segundo plano soluções para manter os que já estão matriculados. Segundo Filho et al. (2007), a evasão representa prejuízos sociais e reflete nas avaliações dos órgãos de fiscalização governamentais. No setor público, significa que os recursos públicos investidos não terão o retorno esperado pelos contribuintes, já no setor privado, é refletido em receita. Além do mais, a evasão torna-se uma fonte de ociosidade de professores, funcionários, equipamentos e espaço físico (LOBO *et al.*, 2009). Para os alunos, a evasão representa a perda de oportunidades de trabalho, crescimento pessoal, melhoria de renda, entre outras consequências.

Conforme Tontini e Walter (2014), embora existam estudos que abordem variáveis que levam os alunos a evadirem, são encontradas poucas pesquisas na literatura nacional que tenham como objetivo identificar métodos que auxiliem as IESs no diagnóstico do real risco de evasão. Segundo Silva (2013), as metodologias encontradas na maior parte dos trabalhos nacionais tratam da evasão em casos específicos, impedindo que os resultados sejam replicados em outras IESs.

Com a difusão do uso de sistemas informatizados nas universidades, o volume de dados gerados e armazenados em bases de dados vem crescendo diariamente (RIGO et al., 2014). Entretanto, o acesso a bases de dados torna-se pouco útil sem o auxílio de ferramentas adequadas para análise e interpretação dos dados, a fim de que professores e gestores possam extrair informações úteis para tomadas de decisão (GOTTARDO et al., 2014). Técnicas de mineração de dados, através do processo de descoberta de conhecimento, podem ser utilizadas para a extração de informações úteis, visando gerar modelos de apoio aos gestores universitários, permitindo ações baseadas no diagnóstico precoce dos fatores relacionados ao fenômeno da evasão (RIGO et al., 2014; MANHÃES et al., 2012).

O objetivo deste artigo é através de uma pesquisa bibliográfica exploratória, apresentar trabalhos que abordem a mineração de dados como ferramenta de apoio ao combate da evasão escolar em instituições de ensino superior brasileiras, além disso, são destacadas técnicas de mineração, atributos, acurácia e metodologia de cada trabalho. Espera-se que este artigo possa ajudar a orientar futuros trabalhos sobre o tema.

## 2. EVASÃO ESCOLAR

Existem diversos conceitos para a evasão escolar, dentre eles, Baggi e Lopes (2011) definem como a saída do aluno da instituição antes da conclusão do curso. A evasão é uma situação complexa, no qual é consequência de um conjunto de fatores que pesam na decisão do aluno em continuar ou não no curso.

Em 1995, a Secretaria de Educação Superior (SESU) do estado de São Paulo, juntamente com o Ministério da Educação, promoveram o Seminário “Evasão nas Escolas Públicas Brasileiras”, dando o ponto de partida para os estudos da evasão em IESs no Brasil. Para analisar o fenômeno da evasão, criou-se a “Comissão Especial de Estudo sobre Evasão” por meio de uma única metodologia para se encontrar soluções para os índices de evasão (POLYDORO, 2000). Silva Filho (2007) define duas categorias de evasão, são elas: (i) evasão anual: aponta a diferença entre alunos matriculados de um ano para outro, assim, se uma IES tem 100 alunos matriculados em um determinado curso e apenas 80 renovam suas matrículas, a evasão anual média no curso seria de 20%; e (ii) evasão total: mostra o número de alunos matriculados em comparação ao número de alunos que concluem o curso, nesse caso, se 100 estudantes entram em um curso e só 54% terminam, a evasão é de 46%.

Apesar da falta de recursos financeiros dos alunos que muitas vezes é apontada como o principal motivo de evasões, essa afirmação é muito simplista. Gisi (2006) acredita que nos setores sociais menos favorecidos, a permanência é ameaçada não apenas pela falta de recurso, mas também pela falta de bagagem cultural e educacional do aluno ingressante. Grande parte do alunado ingressante é fruto de uma escola pública defasada. Logo, a falta de oportunidades em relação a um conhecimento mais amplo provoca dificuldades que desfavorecem sua permanência no ensino superior. Dessa maneira, o apoio que a instituição disponibiliza ao aluno pode ser um diferencial. Silva (2013) aponta quatro conjuntos de fatores que contribuem para a evasão: (i) peculiaridades prévias ao ingresso do estudante na IES como aspectos familiares, habilidades e escolaridade; (ii) a relação entre os objetivos e comprometimento da instituição e dos alunos; (iii) as relações estabelecidas no ambiente acadêmico e social, como performance acadêmica, relacionamento com funcionários, professores e colegas, e atividades extracurriculares; e, (iv) integração acadêmica e a integração social que os itens anteriores proporcionam.

### 3. CIÊNCIA DE DADOS

O elemento central na ciência de dados é o conjunto de dados utilizado para treinamento de algoritmos. O desafio de analisar e interpretar grandes volumes de dados abre a necessidade de novas perspectivas e oportunidades de avanços tecnológicos relacionados ao desenvolvimento de técnicas, metodologias, modelos, algoritmos e arquiteturas (ZIVIANI, 2015). Neste sentido, o aprendizado de máquina investiga como os computadores podem aprender ou melhorar seu desempenho com base em dados (WITTEN, 2000). Para Witten (2000), o aprendizado de máquina permite que programas de computador aprendam automaticamente a reconhecer padrões complexos e a tomar decisões inteligentes com base nos dados.

O campo de aprendizagem de máquina pode ser classificado em três categorias principais, a saber são: (i) **aprendizagem supervisionada**: o conjunto de dados de treinamento possui dados rotulados por um especialista, que são utilizados para supervisionar o aprendizado do modelo de classificação; (ii) **aprendizagem não supervisionada**: os dados de entrada não são rotulados como classes, e, através de técnicas como *clustering*, novas classes de dados são descobertas; e (iii) **aprendizagem por reforço**: através de recompensas que podem ser positivas ou negativas, um agente computacional aprende em um ambiente desconhecido com agir e tomar decisões (RUSSEL, NORVIG 2002).

#### 3.1 Mineração de Dados

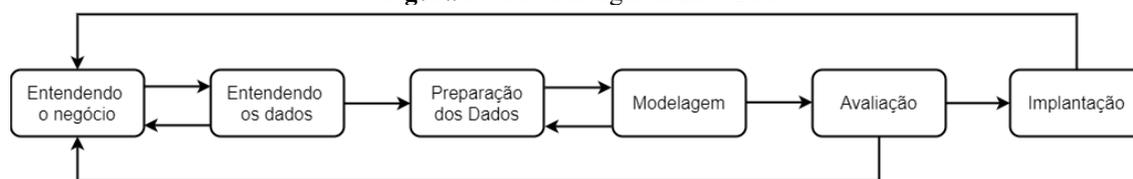
Hand, Mannila e Smyth (2001) definem métodos de mineração de dados como um domínio altamente orientado a aplicativos, onde foram incorporadas diversas técnicas pertencentes a outros domínios, como: estatística, aprendizado de máquina, reconhecimento de padrões, sistemas de banco de dados, dados retrospectivos, visualizações, algoritmos, computação de alto desempenho, entre outros. Neste sentido, os autores definem que o sucesso de aplicações de mineração de dados se dá pela sua interdisciplinaridade.

De acordo com Amo (2003), as tarefas mais conhecidas de mineração de dados são: análises de regras de associação, análise de padrões sequenciais, classificação e predição, análise de agrupamentos, e análise de outliers. Ainda o mesmo autor defende que as técnicas de mineração consistem na especificidade de métodos que garantam como descobrir padrões relevantes. Dentre as principais técnicas utilizadas em mineração de dados, existem técnicas

estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento-poda-validação.

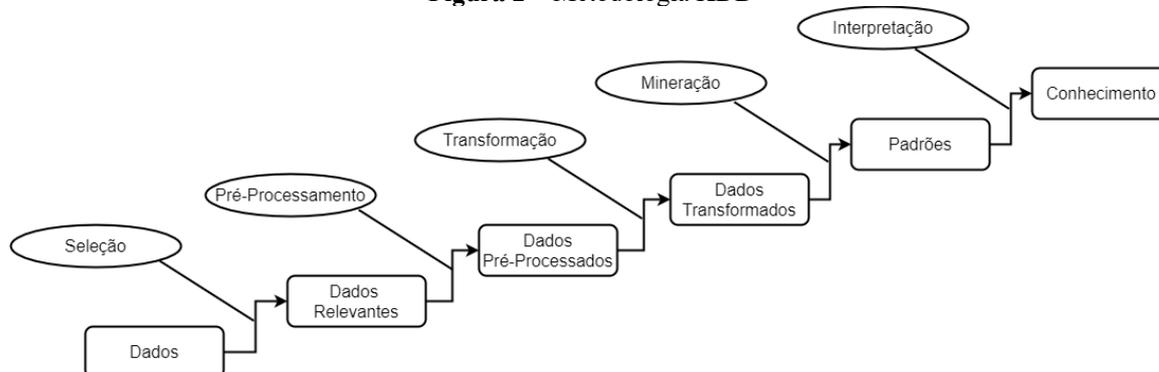
A mineração de dados é um processo criativo que requer variadas habilidades e conhecimentos. Existem alguns frameworks que guiam os processos de mineração de dados, e, dentre os mais utilizados no mercado estão o KDD (do inglês, Knowledge-Discovery in Databases) e CRISP-DM (do inglês, Cross Industry Standard Process for Data Mining). Ambos os frameworks ajudam a traduzir problemas de negócios em tarefas de mineração de dados, além de sugerir transformações de dados e técnicas de mineração de dados apropriadas, e, fornece meios para avaliar a eficácia dos resultados e documentação da experiência (WIRTH, 2000). A Figura 1 ilustra as seis fases do framework CRISP-DM, sendo que as fases não são rígidas, logo, as setas indicam apenas as dependências mais importantes e frequentes entre as fases. Nesta metodologia o processo de mineração não termina depois que uma solução é implantada, o conhecimento adquirido durante o processo pode desencadear novas questões de negócio (FAYYAD, 1996). Já a Figura 2, apresenta o framework KDD, no qual é composto por seis fases. Esta metodologia não foca em questões de negócio ou geração de modelos, mas sim em descoberta de conhecimento a partir de dados.

**Figura 1 – Metodologia CRISP-DM**



Adaptado de (Wirth, 2000)

**Figura 2 – Metodologia KDD**



Fonte: Adaptado de (Fayyad et al, 1996)

### 3.2 Mineração de Dados aplicado a evasão escolar

Ali (2013) aponta em sua pesquisa várias maneiras de empregar mineração de dados em setores educacionais, sendo algumas delas com foco no combate à evasão escolar, são elas: (i) previsão de matrícula de alunos: através de modelos de classificação e regressão, são identificados alunos com risco de evasão baseados na pré-matrícula; (ii) desenvolvimento de currículo: através de árvores de decisão e floresta randômica, são feitas correlações entre curso, categoria e profissões; (iii) direcionamento de alunos: através de classificação bayesiana e árvore de decisão, alunos são segmentados a fim de criar estratégias individualizadas; (iv) seleção de curso: através de redes neurais, são identificadas as satisfações dos alunos com o curso; e (v) desistência de alunos: através do modelo de rede de markov, são encontrados diferentes grupos de alunos associados à diferentes graus de risco de evasão.

## 4. PLANEJAMENTO DA PESQUISA

Esta seção descreve a metodologia de busca de trabalhos relacionados a mineração de dados no contexto de evasão escolar. As etapas de busca dos trabalhos adotadas nesta pesquisa foram: (i) definição de questões de pesquisa; (ii) conduzir a busca de artigos relevantes; (iii) triagem de trabalhos; (iv) redação dos resumos; e (v) extração e mapeamento de dados.

### 4.1 Questões de Pesquisa

O objetivo principal desta pesquisa é identificar estudos primários que relatem o uso de mineração de dados no combate à evasão escolar em instituições de ensino superior brasileiras. Esta pesquisa também visa obter uma visão abrangente sobre as principais técnicas e metodologias de ciência de dados empregadas neste contexto. Para isso, foram estabelecidas as seguintes questões de pesquisa (QP):

- **QP1:** *IES brasileiras estão adotando técnicas de ciência de dados no combate à evasão escolar?*
- **QP2:** *Quais modelos e metodologias estão sendo empregadas na mineração de dados no contexto do combate à evasão escolar no ensino superior?*

### 4.2 Procedimento de Pesquisa

Considerando as questões de pesquisa, foram definidas estratégias de busca, que compreende: critérios de seleção de fontes, lista de fontes, idioma de estudos e palavras-chave com seus termos relacionados:

- **Crítérios de seleção das fontes:** foram adotados os seguintes critérios para selecionar as fontes de busca: atualização de conteúdo (as publicações são atualizadas regularmente); disponibilidade (o texto completo dos artigos está disponível); qualidade dos resultados (precisão dos resultados devolvidos pela pesquisa); e base dados aberta (a base de dados oferece trabalhos gratuitos a comunidade);
- **Base de dados:** de acordo com os critérios de busca, foi adotado como motor de busca o portal Google Acadêmico<sup>1</sup>;
- **Idioma dos estudos:** Somente estudos primários escritos em inglês ou português serão considerados nesta pesquisa;
- **String de pesquisa:** foi utilizado o operador booleano **OU** para vincular os termos relacionados. Todos esses termos foram combinados usando o operador booleano **E** para compor a *string* de pesquisa, conforme exibido na Quadro 1. Além da *string* de pesquisa foram definidos aos anos de publicação 2016, 2017, 2018, 2019, e 2020.

**Quadro 1 - String de Busca**

---

("data mining" E "evasão escolar") OU ("mineração" E "evasão escolar") OU (mineração de dados" E "instituição educacional")

---

Fonte: Autoria própria

#### 4.3 Critérios de inclusão e exclusão de trabalhos

Para seleção dos trabalhos foram adotados Critérios de Inclusão (CI) e Critérios de Exclusão (CE). Esses critérios permitem incluir estudos primários relevantes para responder às questões de pesquisa e excluir estudos que não as respondam. Um estudo primário é incluído se satisfizer um ou mais CIs e excluído se satisfizer um ou mais CEs. Assim, os critérios de inclusão adotados são

- **CI1:** Trabalho aborda técnicas de mineração de dados como ferramenta ao combate à evasão escolar;
- **CI2:** A evasão escolar está no contexto de instituições de ensino superior brasileiras;
- **CI3:** Os trabalhos abordam técnicas, modelos, fonte da base de dados e acurácia da técnica adotada na mineração de dados.

Os critérios de exclusão adotados são:

- **CE1:** O trabalho não aborda estudos primários no contexto de mineração de dados para ao combate à evasão escolar;

---

<sup>1</sup> <http://scholar.google.com.br>

- **CE2:** O estudo principal é um editorial, opinião, tutorial, pôster ou painel;
- **CE3:** O trabalho foi escrito em idioma que não seja português ou inglês;
- **CE4:** O estudo principal tem apenas um resumo ou não está disponível na íntegra.

## 5. RESULTADOS E DISCUSSÕES

A consulta dos trabalhos foi efetuado no período de 14/09/2020 a 15/09/2020 na ferramenta de busca Google Acadêmico, no qual foi utilizando a *string* de pesquisa abordados na **Quadro 1**. A consulta retornou um total de 154 trabalhos e após a aplicação dos critérios de inclusão e exclusão abordados na Seção 4.3, foram selecionados 13 trabalhos conforme exibido na **Quadro 2**.

**Quadro 2 – Trabalhos resultantes**

Autor	Técnica	Variável Relevante	Acurácia	Metodologia	Base de Dados
BARRETO et al (2019)	Algoritmo J48 (árvore de decisão)	Idade do aluno, renda familiar e tipo de curso (tecnologia)	74%	----- ----	Sistema acadêmico próprio
PAZ e CAZELLA (2017)	Algoritmo J48 (árvore de decisão)	Etapa semestral, região da estadia dos alunos e bolsas	90%	DCBD	Coleta de dados próprios (2016)
LANES e ALCÂNTARA (2018)	Algoritmo J48 (árvore de decisão)	Incentivos fornecidos (bolsas), faixa de idade e área do curso	90,7%	----- ----	Sistema acadêmico próprio
RABELO et al (2017)	Clusterização, algoritmos J48 e ID3 (árvore de decisão)	Ações de visualização do curso, adição de postagem em fórum, e envios de tarefas.	Entre 93.9% e 96.5%	----- ----	Própria base de dados
JUNIOR et al (2019)	Árvore de decisão, Floresta randômica, CHAID, stump de decisão, ID3, Indução de regras e Indução de regra única	Faltas e Notas	78,38% e 68,42%	KDD, CRISP	Sistema acadêmico próprio
BARRETO et al (2019)	Algoritmo J48 (árvore de decisão)	Jovens com até 21 anos de idade e maiores que 27 e renda familiar de até dois salários mínimos	79%	KDD	Sistema acadêmico próprio (2017 e 2018)
JUNIOR et al (2018)	baseados em regras (JRip), árvore de decisão (J48) e modelagem estatística (Naive Bayes)	Coefficiente de regressão, dificuldade nas disciplinas, coeficiente de rendimento, percentual de aprovação, total semestre truncados	83% a 87%	----- ----	Própria base de dados
PIMENTEL et al (2019)	NNLM Feedforward	Nota e Bilhete atrasado	84,6%	SS- DetChurn	CRM próprio

Soares et al (2020)	Floresta aleatória e Validação cruzada	Cadastro de alunos, matrizes curriculares e histórico acadêmico	97,4%	AM	12 cursos presenciais da Unitins
GONLÇALVES e BELTRAME (2020)	CN2 indutor de regras, Floresta aleatória, SVM, Regressão logística, Naive Bayes, AdaBoost, Rede neural, SGD e empilhamento	Frequência, locomoção e moradia dos alunos	90,5%	KNN	Banco de dados – IFES (2014 a 2018)
SILVA e CRUZ (2019)	Redes neurais	Curso, ano letivo, disciplinas, notas, classificação no exame de seleção, tipo de concorrência, idade e município de origem	80% a 93%	KDD	Sistema acadêmico – IFAL Magogi (2015 a 2017)
MARTINS et al (2020)	Florestas randômicas, SVM e RNA	Matrículas demográficas, socioeconômicas e de acesso ao ensino superior	86,30% a 88,96%	----- ----	Matrículas registradas da base do IPB
FUNCHAL et al (2016)	árvores de decisão com os algoritmos C4.5 e J48	Reprovações, aprovações, média, português, literatura, física, biologia, matemática e química	83,9% a 99,3%	KDD	Dados dos alunos de engenharia de computação da FURG

Fonte: Autoria própria

Nas fases de aplicação dos critérios de inclusão e exclusão, além dos trabalhos de que abordam o ensino superior, observou-se numerosa quantidade de publicações de trabalhos que utilizam a ciência de dados no combate à evasão escolar no contexto do ensino fundamental e médio. Uma das variáveis que explica este grande interesse são as bases de dados governamentais que estão disponíveis publicamente referentes ao ensino fundamental e médio.

Os trabalhos desenvolvidos no contexto do ensino superior são caracterizados principalmente por utilizarem na grande maioria bases de dados de sistemas próprios como por exemplos sistemas acadêmicos. Isso reflete diretamente nas variáveis abordadas nos trabalhos, uma vez que características com relevâncias pessoais e sociais não são apontadas e talvez sejam pertinentes neste tipo de estudo, como por exemplo, se o aluno estudou em escola pública, se mantém atividades de trabalho em paralelo aos estudos, se possui filhos menores de idade, entre outros.

Nota-se que a técnica de aprendizagem de máquina mais utilizada no cenário de evasão escolar é a árvore de decisão, no qual totaliza 53,33% do total dos trabalhos selecionados, sendo que estes trabalhos têm média de acurácia acima dos 86%, o que demonstra eficiência na extração de conhecimento.

Em relação a metodologia utilizada na extração de conhecimento, 38% dos trabalhos selecionados não citam qual a metodologia foi utilizada. Essa informação é essencial para que

os resultados alcançados nos trabalhos possam ser replicados com exatidão em trabalhos futuros, portanto, é necessário que os próximos trabalhos tornem explícito a metodologia utilizada.

## 6. CONSIDERAÇÕES FINAIS

A evasão escolar é um assunto alarmante para as instituições de ensino superior brasileiras e é protagonista da análise que o presente trabalho dispõe em conjunto com a ciência de dados. Em vista de que a maioria das instituições prezam pela captação de alunos, deixando em segundo plano soluções para manter os que já estão matriculados, a pesquisa teve base em trabalhos que evidenciam soluções de mineração de dados para análise e melhoria da evasão escolar no ensino superior brasileiro.

Este trabalho demonstra que a ciência de dados pode auxiliar no combate à evasão nas instituições de ensino superior brasileiras com grande assertividade utilizando técnicas de mineração de dados. Como propostas para trabalhos futuros, pode-se analisar e comparar a técnica de árvore de decisão com as demais, através de estudos de casos afim de extrair qual delas pode apresentar melhor eficiência. Além disso, trabalhos podem ser elaborados no sentido de buscar informações sociais e pessoais dos discentes, para assim buscar acurácias próximas a 100%. Por fim, é recomendável que os próximos trabalhos tornem explícito quais metodologias de ciência de dados foram utilizadas, para que trabalhos futuros possam replicar as técnicas abordadas.

## REFERÊNCIAS

Amo, S. **Técnicas de Mineração de dados**, Universidade Federal de Uberlândia, 2003, disponível em <https://bit.ly/2ScEQAh>.

Ali, M. Role of **Data Mining in Education Sector**. International Journal of Computer Science and Mobile Computing Vol.2 Issue. 4, April- 2013, pg. 374-383

Baggi, C.; LOPES, D. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. **Avaliação** – Campinas, julho, v. 16, n. 2, p. 355-374, 2011.

Barreto, D. et al **Evasão no ensino superior: Investigação das causas via mineração de dados**, Educação Profissional e Tecnológica, Vol.3, No.2, dezembro 2019, disponível em <https://bit.ly/2SfIVol>.

Fayyad, U, et al. **From Data Mining to Knowledge Discovery in Databases**, AI Magazine, Vol.17, No.3, março 1996, disponível em <https://bit.ly/3kVXCro>.

Funchal, J. et al. **Um Modelo Preditivo para Estudo da Evasão na Graduação Utilizando Mineração de dados**, Revista de Tecnologia, Vol.9, No.3, dezembro 2016, disponível em <https://bit.ly/34h7oy2>.

Filho, R. et al. **A evasão no ensino superior brasileiro. Cadernos de Pesquisa**. Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia. v. 37, n. 132, p. 641-659, set./dezembro 2007, disponível em <https://bit.ly/2SbyH7g>.

Gisi, M. **A Educação Superior no Brasil e o caráter de desigualdade do acesso e da permanência**. Diálogo Educacional, Curitiba, v. 6, n. 17, p. 97-112, 2006.

Gonçalves, O; Beltrame, W. **Mineração de dados e Evasão estudantil: Analisando o curso de nível superior do IFES**, XII Congresso de Administração, Sociedade e Inovação, dezembro 2019, disponível em <https://bit.ly/30kwM4W>.

Gottardo, E.; et al. **Estimativa de desempenho acadêmico de estudantes: análise da aplicação de técnicas de mineração de dados em cursos à distância**. Revista Brasileira de Informática na Educação, v. 22, n. 1, p. 45-55, 2014.

Han, J; Kamber, M; Pei. J. **Data Mining: Concepts and Techniques, Third Edition**. University of Illinois at Urbana-Champaign, 2011.

Hand, D; Mannila, H.; Smyth, P. **Principles of Data Mining**. Massachusetts Institute of Technology, 2001.

Júnior, J. et al. **Método de Seleção de Atributos Aplicados na Previsão da Evasão de Cursos de Graduação**, Revista de informática aplicada, Vol.13, No.2, fevereiro 2017, disponível em <https://bit.ly/30iRVfG>.

Junior, R. et al. **Descoberta de Conhecimento para Identificação de Fatores que Influenciam o Desempenho Discente**, Informática na Educação: teoria & prática, Vol.22, No.3, dezembro 2019, disponível em <https://bit.ly/36hEMYg>.

Lima, M; Fagundes, R. **Education Data Mining: A Study of the Factors That Cause School Dropout in Higher Education Institutions in Brazil**, Revista Novas Tecnologias na Educação, Vol.18, No.1, julho 2020, disponível em <https://bit.ly/3jjbHz9>.

Lanes, M; Alcântara, C. **Predição de Alunos com Risco de Evasão: estudo de usando mineração de dados**, VII Congresso Brasileiro de Informática na Educação, outubro 2018, disponível em <https://bit.ly/3kZ9s4D>.

Lobo, E. **A evasão no ensino superior brasileiro: novos dados**. Blog Estadão, São Paulo, 2017. Disponível em: <https://bit.ly/30n7bs3>. Acesso em: 24 ago. 2020.

Lobo, R.L. et al. **Evasão no ensino superior: causas e remédios** - Jun. 2009. Disponível em: <https://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>. Acesso em: 24 ago. 2020.

Mannila, H. **Methods and Problems in Data Mining**, In Proc. 7<sup>th</sup> International Conference on Database Theory (ICDT 97), Delphi, Greece, 1997, disponível em <https://bit.ly/34aPD3v>.

Manhães, L. M. B. et al. **Identificação dos Fatores que Influenciam a Evasão em Cursos de Graduação Através de Sistemas Baseados em Mineração de Dados: Uma Abordagem Quantitativa**. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012), São Paulo, 2012

Martins, M. et al. **Previsão do Abandono Acadêmico numa instituição de ensino superior com recursos a data mining**, Revista Ibérica de Sistemas e Tecnologias de Informação, Vol.28, No.4, abril 2020, disponível em <https://bit.ly/3iisnWa>.

Pimentel, T. et al **Mineração de Padrões Sequenciais de Sentimentos: Um Estudo de Caso na Prevenção de Evasão da Educação Superior**, VIII Congresso Brasileiro de informática na educação, nov. 2019, disponível em <https://bit.ly/3kZ6RaS>.

Polydoro, S. **O trancamento de matrícula na trajetória acadêmica do universitário: condições de saída e de retorno a instituição**. 2000. 175f. Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Educação, Campinas, SP. Disponível em <https://bit.ly/2S9TYhU>.

Paz, F; Cazella, S. **Identificando o perfil de evasão de alunos de graduação através da mineração de dados educacionais: um estudo de caso de uma universidade comunitária**, VI Congresso Brasileiro de informática na educação, Nov. 2017, disponível em <https://bit.ly/3cKyv8x>.

Rigo, S. et al. **Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios**. Revista Brasileira de Informática na Educação, v. 22, n. 1, 2014.

Russel, S; Norvig, P. **Artificial intelligence: A modern approach**. Prentice Hall, New Jersey 2002.

Silva, E; Cruz, J. **Mineração de Dados Educacionais: Uso de Redes Neurais Artificiais na Predição do Perfil Acadêmico do Aluno IFAL campus Maragogi**, XIX Escola Regional de Computação Bahia, Alagoas e Sergipe, abril 2019, disponível em <https://bit.ly/30ltvlO>.

Silva, C. et al **Avaliação dos níveis de expectativa e motivação de estudantes de graduação do ensino presencial através da Mineração de dados**, Informática na Educação: teoria & prática, Vol.22, No.3, dezembro 2018, disponível em <https://bit.ly/3l4m5vp>.

Silva, R. et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**. Instituto Lobo para o Desenvolvimento da Educação, Ciência e da Tecnologia. v. 37, n. 132, p. 641-659, set./dez. 2007.

Silva, G. **Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes**. Revista da Avaliação da Educação Superior (Campinas), v. 18, n. 2, p. 311-333, Jul 2013.

Soares, L. et al. **Aplicação de Técnicas de Aprendizado de Máquina em um Contexto Acadêmico com Foco na Identificação dos Alunos Evadidos e não Evadidos**, VII Congresso Brasileiro de Inf. na Educação, nov. 2020, disponível em: <https://bit.ly/2G3j2Vt>.

Tontini, G.; Walter, S. **Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior**. Revista da Avaliação da Educação Superior (Campinas), v. 19, n. 1, p. 89-110, mar 2014.

Wirth, R.; Hipp, Jochen. **CRISP-DM: Towards a standard process model for data mining**. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining. 2000.

WITTEN, Ian, et al. **Data Mining: Pratical Machine Learning Tools and Techniques with java implementations**, San Francisco: Morgan Kaufmann Publisher, 2000.

Ziviani, A. et al. **Ciência de dados: desafio para a ciência, indústria e governo**. Revista Eletrônica de Jornalismo Científico, setembro 2015, disponível em <https://bit.ly/34dnHvR>.