



## A COMBINAÇÃO DE TECNOLOGIAS NO COMBATE À EVASÃO ESCOLAR NA ÁREA DE INFORMÁTICA

**Átila Pires dos Santos** – atila.santos@ifb.edu.br  
Instituto Federal de Brasília (IFB)  
QNM 40, Área Especial nº 01  
72146-000 – Taguatinga – DF

**Sandra Isaelle Figueiredo dos Santos** – sandraisaelle@gmail.com  
Universidade de Brasília  
Campus Universitário Darcy Ribeiro  
70910-900 – Brasília – DF

**Vandor Roberto Vilardi Rissoli** – vandorissoli@gmail.com  
Universidade de Brasília  
Campus Universitário Darcy Ribeiro  
70910-900 – Brasília – DF

**Resumo:** *Os cursos de nível superior na área de Informática vêm sofrendo o problema da evasão discente. Propostas diferentes para o combate dessa evasão vêm sendo elaboradas, sendo por meio deste trabalho apresentada a combinação das tecnologias de Mineração de Dados e Sistemas Tutores Inteligentes no enfrentamento desta realidade. Essa combinação propicia uma predição resultante da inferência sobre o acompanhamento contínuo da realidade de cada estudante, podendo ser solicitada em tempo real pelo docente que ainda receberá apoio significativo para a definição de sua estratégia pedagógica personalizada com a realidade de cada um de seus aprendizes.*

**Palavras-chave:** *Evasão Escolar, Mineração de Dados, STI, Cluster, TI na Educação*

### 1. INTRODUÇÃO

Os cursos de nível superior na área de Informática (Ciência da Computação, Sistemas de Informação, Engenharia de Software) vêm sofrendo com o problema da evasão discente. Embora este problema se estenda a todas as graduações em engenharia (ALI; SHUBRA, 2010) e até mesmo graduações de outras áreas, Silva et al. (2012, p. 1) alertam sobre a gravidade deste problema para os cursos de nível superior na área de Informática: “as altas taxas de evasão em cursos de graduação atingem, em alguns casos extremos, valores superiores a 80%”.

Quando se torna possível antever um problema, a intervenção sobre ele pode acontecer. Atualmente, várias são as propostas que procuram tratar essa evasão, estando entre



elas a predição apoiada por técnicas de mineração de dados. Esta abordagem, usualmente, tem resultados mais promissores quando utiliza dados que são atualizados com frequência. Em geral, os dados dos discentes são atualizados apenas uma vez a cada período letivo (ou seja, uma vez a cada semestre, na maioria dos cursos de graduação).

Santos et al. (2013) apresentam a proposta de realizar a predição da evasão valendo-se de dados fornecidos por um Sistema Tutor Inteligente (STI), conhecido como SAE (Sistema de Apoio Educacional). Estes dados, constantemente fornecidos pelos atores atuantes no SAE (aluno, professor, etc.), são obtidos em tempo real. No entanto, a qualidade dos dados desta abordagem proposta por estes autores dependem, diretamente, de quanto o docente incentiva seus alunos ao uso ativo e contínuo do SAE.

Desta forma, questiona-se: considerando-se as múltiplas dimensões (variáveis) analisadas, o quanto um professor precisa incentivar seus estudantes a trabalharem, de maneira mais ativa, como se propõe este STI, para que seja possível predizer a evasão?

## 2. REFERENCIAL TEÓRICO

O primeiro passo para compreender o problema da evasão é buscar por uma definição desse fenômeno. Dore e Lüscher (2011, p. 775) discorrem sobre o problema da definição do termo evasão, tendo este sido associado a diversas situações, como “a retenção e repetência do aluno na escola, a saída do aluno da instituição, sua saída do sistema de ensino, a não conclusão de um determinado nível de ensino, o abandono da escola e posterior retorno” e também “àqueles indivíduos que nunca ingressaram em um determinado nível de ensino, especialmente na educação compulsória, e ao estudante que concluiu um determinado nível de ensino, mas se comporta como um *dropout*”, ou seja, como um aluno evadido. Desta maneira, torna-se difícil desassociar a evasão do sucesso ou insucesso escolar.

Para seu estudo, Martins (2007, p. 29) definiu a evasão como “a saída do aluno de uma IES (Instituição de Ensino Superior) ou de um de seus cursos de forma temporária ou definitiva por qualquer motivo, exceto a diplomação”. De maneira semelhante, Gaioso (2005, p. 9) define a evasão como sendo a “interrupção no ciclo de estudo”. Na pesquisa elaborada por esta autora (p. 38), considerou-se o aluno como evadido quando o mesmo “deixou o curso por qualquer motivo que não seja a obtenção da titulação”.

Martins (2007, p. 29) sintetiza a questão abordada pela Comissão Especial de Estudos sobre Evasão nas IES Públicas (BRASIL/MEC, 1997) sobre a evasão nos diferentes *locus* do nível superior, ou seja, caracteriza três estágios da evasão no nível superior, sendo eles: 1) evasão de curso: “quando o estudante desliga-se do curso superior em situações diversas tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou reopção (mudança de curso), trancamento, exclusão por norma institucional”; 2) evasão da instituição: “quando o estudante desliga-se da instituição na qual está matriculado”; e 3) evasão do sistema: “quando o estudante abandona de forma definitiva ou temporária o ensino superior”. Existe uma sequência de pré-requisitos entre estes três níveis de evasão: para que a evasão da instituição aconteça, é necessário evadir também do curso ou programa. Da mesma forma, para haver a evasão do sistema, o aluno também deverá evadir da instituição, que implica na evasão do curso ou programa.

## 2.1. Mineração de Dados

Witten e Frank (2005, p. 5) definem mineração de dados como o “processo de descobrimento de padrões em dados”. Dekker et al. (2009, p. 42) consideram o uso de técnicas de mineração de dados para predição da evasão discente como sendo “relativamente nova”. Durante o levantamento de trabalhos relevantes sobre este tema foi confirmada tal afirmação, pois nenhuma publicação foi encontrada antes do ano 2000. Além disso, grande parte da bibliografia utilizada por estes trabalhos, como referencial teórico, é composta por conteúdos que não tratam diretamente o problema da evasão discente, tendo como objeto de estudo o sucesso ou insucesso escolar.

A abordagem usada por estes artigos é semelhante, variando, principalmente, no quesito de quais atributos foram trabalhados e qual ou quais algoritmos de mineração de dados foram aplicados. Os atributos dos estudantes mais frequentemente encontrados nesses estudos foram: situação/notas do aluno em disciplinas já cursadas (incluindo aqui o total de reprovações que ele já teve e sua frequência), sexo, estado civil, idade, profissão, renda familiar, endereço residencial e colégio onde cursou seu ensino médio.

A escolha dessas variáveis é justificada por Cobbe et al. (2011) pelo levantamento feito por Gaioso (2005), que apresenta como algumas das principais causas da evasão do nível superior: a necessidade de trabalhar / horário de trabalho incompatível com o de estudo, problemas financeiros, casamento / nascimento de filhos, desconhecimento da metodologia do curso escolhido, deficiência da educação básica e reprovações sucessivas.

Vale ressaltar que há ainda outras variáveis, presentes em apenas alguns estudos, como: nota do aluno no processo seletivo de admissão na IES e o nível de escolaridade do pai e/ou mãe do estudante. Superby et al. (2006) vão além, trazendo como variável em seu estudo a estrutura familiar do aprendiz e até mesmo o fato do aluno ser fumante.

Quanto à escolha do algoritmo de mineração de dados, a maioria dos trabalhos pesquisados utiliza mais do que um tipo e os compara e/ou os agrega em um comitê. Dekker et al. (2009) concluíram que, possivelmente, os algoritmos de classificação por árvore de decisão sejam as melhores escolhas e que os algoritmos de *clustering* podem auxiliar na criação das classes e categorização dos estudantes dentro destas classes, como feito por Campello e Lins (2008).

## 2.2. O Sistema Tutor Inteligente SAE

No Sistema de Apoio Educacional (SAE) cada estudante interage com os recursos fornecidos por este STI, classificado como ITA (*Intelligent Teaching Assistant*), que usa a Lógica *Fuzzy* em seus processos de inferência para modelar os aspectos cognitivos de cada aprendiz. Esse assistente inteligente (SAE) serve de tutor virtual aos aprendizes durante todo o seu período letivo e no estudo sobre o conteúdo a ser assimilado, por meio da observação do comportamento interativo de cada um na realização de exercícios e tarefas, contatos com os monitores e a participação de atividades colaborativas com os demais estudantes de sua turma, além de seu próprio professor.

No ambiente virtual do SAE cada aprendiz pode acompanhar seu próprio esforço, participação e desempenho alcançado em cada conteúdo de aprendizagem, sendo orientado, por este ITA, a possível necessidade de maior atenção e mais tempo de dedicação ao estudo sobre os tópicos específicos de um conteúdo em que estão sendo detectadas deficiências momentâneas de assimilação, podendo estas comprometerem o sucesso de sua aprendizagem.

As análises e as variáveis linguísticas utilizadas pelo SAE, sobre a realidade de cada aprendiz, são explicitadas por Rissoli e Santos (2011), sendo seu principal foco a averiguação dos termos linguísticos mais coerentes as variáveis: **participação**, **esforço** e **desempenho** do aprendiz. A primeira variável, **participação**, infere “a participação de cada aprendiz nas atividades interativas propostas pelo docente”, que “poderão acontecer por meio de fóruns ou chats (bate-papo) envolvendo assuntos relacionados aos conceitos pertinentes a cada conteúdo”. A variável **esforço** averigua “o número de exercícios resolvidos e a quantidade de visitas que cada estudante efetuou na monitoria estudantil” por tópico ou conceito abrangido por cada conteúdo, enquanto que a variável **desempenho** “envolve o resultado obtido na solução da quantidade de exercícios apurados pela variável esforço” ou pelas tarefas solicitadas pelo próprio SAE e pelo docente (RISSOLI; SANTOS, 2011, p. 7-9).

Rissoli (2007, p.30) ressalta que “estas dificuldades podem contribuir ainda mais com a falta de motivação do estudante” e que isto pode promover o “acúmulo de conteúdo a ser estudado, timidez em expressar suas dúvidas em sala de aula, principalmente pela dificuldade em formular suas questões, além da não participação efetiva nos trabalhos elaborados em grupo”, e que, desta maneira, poderia contribuir com “a evasão escolar”. Sendo assim, por meio do SAE e de informações sobre o contexto do aprendiz dentro e fora da sala de aula, é possível prever, com maior segurança, um quadro de possível evasão em tempo real, possibilitando uma ação de maior intervenção docente e do próprio ITA junto a cada estudante que o utilize como recurso de apoio educacional.

### 3. METODOLOGIA

Para este estudo adota-se o conceito de evasão proposto por Gaioso (2005, p. 38) e corroborado por Martins (2007), já apresentado anteriormente: a evasão ocorre quando o estudante “deixou o curso por qualquer motivo que não seja a obtenção da titulação”, sendo que a evasão não necessariamente possui caráter permanente. Adota-se ainda os três *locus* de evasão propostos pela Comissão Especial (BRASIL/MEC, 1997), ou seja, o curso, a instituição e o sistema de ensino. Acrescentam-se neste trabalho mais dois atributos que são o semestre (ou módulo) em curso e a própria disciplina de estudo onde o aluno está matriculado.

Desta forma, o estudante evadido, ou em processo de evasão, realiza uma série de abandonos, em cinco etapas: 1) abandono da disciplina que está cursando; 2) abandono do semestre ou módulo; 3) desistência do curso; 4) saída da instituição (IES); 5) abandono do nível superior como um todo. Depreende-se que a segunda etapa tem como pré-requisito a primeira (seja ela ao longo de um período letivo, seja ela entre um período letivo e outro). Da mesma forma, para alcançar a terceira etapa, ele deverá ter realizado a segunda, que implica ter realizado a primeira. O mesmo vale para a quarta e quinta etapas. Assim, um estudante que alcance as últimas etapas de abandono deverá ter passado antes pela primeira etapa.

A primeira fase deste experimento foi o de pré-processamento, onde foram selecionados atributos dos alunos. A escolha destes atributos se baseia nas escolhas de estudos anteriores e na disponibilidade de dados de alunos contidos no SAE. Desta forma, os dados que foram selecionados são: *i*) total de questões obrigatórias respondidas pelo aluno (listas de exercícios solicitadas pelo docente); *ii*) total de questões obrigatórias respondidas corretamente pelo aluno; *iii*) total de questões avulsas respondidas pelo aluno (realizadas de



maneira proativa pelo estudante); *iv*) total de questões avulsas respondidas corretamente pelo aluno; *v*) número de acessos realizados ao SAE pelo aluno; *vi*) nome da disciplina; *vii*) nome do professor; *viii*) idade do aluno e *ix*) seu sexo. As variáveis de *i* a *v* foram usadas para a divisão do universo de alunos em *clusters*, como foi feito em Santos et al. (2013), e as outras variáveis foram utilizadas para validação dos resultados.

Utilizou-se neste estudo o software de mineração de dados WEKA (*Waikato Environment for Knowledge Analysis*), ferramenta genérica e livre para a mineração que suporta diferentes abordagens de aprendizagem de máquina (WASHIO ET AL., 2007). Como observado por Dekker et al. (2009) e executado por Campello e Lins (2008), fez-se uso de *clusters* para agrupar instâncias similares, com o objetivo de criar categorias para classificar a situação de cada aluno. Foi empregado o algoritmo de *clustering* EM (*Expectation-Maximization*) neste estudo, que foi usado com suas configurações padrões, inclusive pelo valor do parâmetro *k*, configurado para sugerir automaticamente o número de *clusters* a ser criado.

Segundo Witten e Frank (2005, p. 137), *k-means* é a “técnica de *clustering* clássica”. Embora seja um método “simples e efetivo”, não há garantias que o melhor resultado alcançado ao final do processo seja o melhor resultado possível para a amostra utilizada, sendo “fácil imaginar situações nas quais o *k-means* não consegue encontrar um bom resultado” (WITTEN; FRANK, 2005, p 137-138). Uma técnica alternativa é o algoritmo EM. Ele é descrito por Witten e Frank (2005, p. 265) como sendo capaz de “convergir para um máximo de maneira garantida”, ainda que este seja um máximo local e não global. Ele calcula as funções de verossimilhança dos *clusters* criados (fase de *Expectation*) e então busca maximizar seus valores (fase de *Maximization*).

A escolha pelo uso de algoritmos de *clustering* se deve a necessidade de classificar as instâncias extraídas do banco de dados do SAE. Por um lado, esta necessidade acontece pelo fato de não haver uma classificação nativa ao SAE acerca da evasão discente, sendo essencial, portanto, que este dado seja fornecido por outra fonte. Por outro lado, independente da fonte consultada, esta classificação estará dividida em apenas duas classes (evadido ou não evadido), não sendo possível assim observar as nuances dentro de cada classe. Esta etapa foi executada sobre a base de dados dos estudantes fornecidas pelo SAE, que inclui 1509 instâncias, cada uma representando um aluno que cursou uma disciplina em um dos dois semestres dos anos de 2010, 2011 ou 2012.

Em seguida, com o objetivo de validar os *clusters* criados na etapa anterior, estes foram comparados à situação real dos alunos, baseando-se em seu total de faltas ao final do semestre de uma disciplina. Todos os alunos com até 25% de faltas foram considerados não-evadidos (NE) nesta disciplina. Alunos com mais de 25% e até 50% de faltas foram classificados como possíveis evadidos (PE) e, acima desta faixa (50% - metade), evadidos (E).

Utilizou-se nesta etapa um recorte do universo total, contendo 786 alunos. Esta amostra de estudantes fornecidos pelo SAE cursavam o Bacharelado em Sistemas de Informação (BSI) ou o Bacharelado em Ciência da Computação (BCC), sendo identificados, respectivamente, neste trabalho como alunos de BSI (total de 442 alunos) e de BCC (total de 344 alunos). Seis docentes, aqui denominados como A, B, C, D, E e F, ministraram aulas para estes estudantes.

#### 4. RESULTADOS

Observou-se que a criação de *clusters* pelo algoritmo EM, sobre a base de dados do SAE, teve como saída vários resultados repetidos. Um resultado demonstrou-se predominante quando permitiu-se ao algoritmo recomendar o valor *k* a ser utilizado. Ao se testar o parâmetro *seed* com os valores de 0 a 49, este resultado foi obtido em 42 (84%) tentativas.

Este resultado fornecido pelo EM contém quatro *clusters*, cada um contendo 744 (49,3%), 319 (21,1%), 303 (20,1%) e 143 (9,5%) instâncias respectivamente. Dos oito resultados divergentes, seis eram idênticos (*seeds* 0, 9, 28, 41, 47 e 48), resultando em três *clusters*, sendo que um deles era o mesmo *cluster* de 744 (49,3%) instâncias. Dos dois restantes, um criou 5 *clusters* (*seed* 24) e o outro, 7 *clusters* (*seed* 18).

Verificou-se que o menor *cluster* (143 instâncias) corresponde aos alunos mais participativos e menos propensos a evadir, seguido pelo segundo menor *cluster* (303 instâncias). O *cluster* de 319 instâncias representa os alunos pouco participativos e o *cluster* de 744 instâncias corresponde aos estudantes muito propensos a evadir ou já evadidos. Este último *cluster* foi dividido em duas partes, sendo que uma delas (115 instâncias) corresponde aos alunos que nunca realizaram nenhuma questão no SAE, seja por meio de listas (obrigatórias) ou avulsa. Estes cinco agrupamentos correspondem aos cinco *clusters* apresentados em Santos et al. (2013): *a*) Alunos Muito Participativos, *b*) Alunos Participativos, *c*) Alunos Pouco Participativos, *d*) Alunos Propensos a Evadir e *e*) Alunos Evadidos.

Dentro do recorte realizado contendo 786 alunos, 77 foram agrupados no *cluster a*, 157 no *cluster b*, 114 no *cluster c*, 387 no *cluster d* e 51 no *cluster e*. A Tabela 1 apresenta a relação entre os cinco *clusters* e os docentes de A a F para o recorte de 786 estudantes.

Tabela 1 - Relação entre *clusters* e professores.

Professor	Cluster a	Cluster b	Cluster c	Cluster d	Cluster e
<b>A</b>	74 (28,14%)	112 (42,59%)	61 (23,19%)	6 (2,28%)	10 (3,80%)
<b>B</b>	3 (1,30%)	40 (17,32%)	21 (9,09%)	150 (64,93%)	17 (7,36%)
<b>C</b>	0 (0,00%)	2 (1,13%)	17 (9,60%)	141 (79,66%)	17 (9,60%)
<b>D</b>	0 (0,00%)	0 (0,00%)	0 (0,00%)	38 (97,44%)	1 (2,56%)
<b>E</b>	0 (0,00%)	0 (0,00%)	10 (19,61%)	36 (70,59%)	5 (9,80%)
<b>F</b>	0 (0,00%)	3 (12,00%)	5 (20,00%)	16 (64,00%)	1 (4,00%)

##### 4.1. Validação dos Resultados

Dos 786 alunos presentes no recorte, 536 são não-evadidos na disciplina (ou seja, têm um total de faltas de 25% ou menos), 138 são possíveis evadidos (entre 25% e até 50% faltas) e 112 são evadidos (acima de 50% faltas).

Do total de discentes validados, 702 são do sexo masculino e 84 do sexo feminino. Destes 702, 485 (69%) não evadiram, 126 (18%) possivelmente evadiram e 91 (13%) evadiram. Das 84 alunas, 51 (60,71%) não evadiram, 12 (14,29%) possivelmente evadiram e 21 (25,00%) evadiram. Mesmo representando apenas 10,69% do total, o sexo feminino parece estar bem representado nas três classes. Já a relação destes 786 alunos validados e suas idades, apresentadas na Tabela 2, parece confirmar, apenas parcialmente, os achados mostrados em Superby et al. (2006) e Santos et al. (2013), ou seja, quanto mais velho é o aluno, maior sua chance de evadir.

Tabela 2 - Relação entre as classes de validação e as idades dos alunos analisados.

Idade	Evadidos	Possivelmente Evadidos	Não Evadidos
<b>Abaixo de 20</b>	9 (8,04%)	17 (12,32%)	91 (16,98%)
<b>20</b>	7 (6,25%)	16 (11,59%)	99 (18,47%)
<b>21</b>	15 (13,39%)	24 (17,39%)	106 (19,78%)
<b>22</b>	22 (19,64%)	25 (18,12%)	86 (16,04%)
<b>23</b>	11 (9,82%)	17 (12,32%)	44 (8,21%)
<b>24</b>	12 (10,71%)	14 (10,14%)	31 (5,78%)
<b>25</b>	7 (6,25%)	6 (4,35%)	16 (2,99%)
<b>Acima de 25</b>	29 (25,89%)	19 (13,77%)	63 (11,75%)
<b>Total</b>	<b>112 (100,00%)</b>	<b>138 (100,00%)</b>	<b>536 (100,00%)</b>

Comparou-se então as três classes elaboradas para validação e os 5 *clusters* criados pelo algoritmo EM para cada professor, no contexto dos 786 alunos validados. Nesta comparação, semelhante a uma matriz de confusão, algumas combinações indicam maior identificação dos *clusters* criados com a realidade e outras combinações indicam menor identificação. Por exemplo, espera-se que todos os alunos do *cluster e* se identifiquem com a classe de evadidos, enquanto todos os alunos do *cluster a* estejam classificados como não evadidos. Quanto mais longe se está deste resultado, menos os *clusters* representam a realidade. Considerou-se aceitáveis para o *cluster d* resultados da classe evadido (E) e possivelmente evadido (PE), e para o *cluster b*, possivelmente evadido e não evadido (NE). O *cluster c* é um grupo intermediário e, portanto, não foram estabelecidos pressupostos de como este grupo deveria se relacionar com a validação.

Esta discrepância entre os *clusters* e os resultados reais de evasão no período parecem confirmar a conclusão de Santos et al. (2013) acerca do não incentivo destes docentes, exceto por dois docentes (A e B), aos seus alunos no uso da ferramenta SAE ao longo da disciplina, causando a falsa impressão de que estes aprendizes haviam evadido. Isto é evidenciado na relação entre o *cluster d* e a classe não evadido (falsos positivos), que conteve quase todos os erros na relação entre *clusters* e classes para o algoritmo EM, exceto para o professor A. Para

os professores C, D, E e F, mais de 50% dos alunos se encontram neste caso específico, como indicado pela Tabela 3.

Tabela 3 - Relação entre os falsos positivos e os professores.

-	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<b>Erro <i>d</i>/NE</b>	2 (0,7605%)	89 (38,5281%)	110 (62,1469%)	35 (89,7436%)	30 (58,8235%)	13 (52,0000%)
<b>Erro <i>e</i>/NE</b>	0 (0,0000%)	0 (0,0000%)	4 (2,2599%)	1 (2,5641%)	1 (1,9608%)	0 (0,0000%)
<b>Erro <i>e</i>/PE</b>	0 (0,0000%)	6 (2,5974%)	6 (3,3898%)	0 (0,0000%)	2 (3,9216%)	0 (0,0000%)
<b>Total Erro</b>	2 (0,7605%)	95 (41,1255%)	120 (67,7966%)	36 (92,3077%)	33 (64,7059%)	13 (52,0000%)
<b>Outros Resultados</b>	261 (99,2395%)	136 (58,8745%)	57 (32,2034%)	3 (7,6923%)	18 (35,2941%)	12 (48,0000%)
<b>Total</b>	263 (100%)	231 (100%)	177 (100%)	39 (100%)	51 (100%)	25 (100%)

Vale ressaltar que casos de falsos negativos, ou seja, alunos evadidos que estavam incluídos nos *clusters a* e *b*, foram encontrados para o professor A (14) e B (1). Isto parece indicar que estes alunos eram participativos antes de evadirem; estes números parecem sugerir casos em que a evasão se deu por outros motivos que não o insucesso escolar, justificando a necessidade da averiguação de outras variáveis além daquelas diretamente ligadas ao rendimento do aluno.

Observou-se que os alunos incluídos pelo algoritmo EM no *cluster d* nunca haviam realizado questões avulsas. Detectou-se ainda que aqueles alunos cuja a soma dos valores das variáveis *i* a *iv* (questões obrigatórias e avulsas) fosse abaixo de 200, em 86% das vezes eles estavam incluídos nos *clusters c*, *d* ou *e*. Como esta soma inclui tanto o total de questões realizadas quanto questões certas e assumindo-se uma proporção mínima de acerto de 50%, recomenda-se aos docentes que incentivem a cada aluno realizar um mínimo de 133 questões para que a predição seja mais confiável e segura.

## 5. CONCLUSÃO

Diante deste estudo é possível concluir que o uso de técnicas de mineração (data-mining) nos dados provenientes do SAE sobre seus aprendizes se mostrou promissor. Por um lado, os dados fornecidos pelo SAE possibilitaram predições sólidas aos algoritmos de mineração. Por outro lado, a criação de *clusters* permitiu a análise dos dados contidos no SAE





de modo que não lhe era possível fazer antes. A mineração de dados foi empregada neste estudo na busca da predição da evasão, mas outros problemas educacionais poderiam ser também trabalhados a partir desta investigação.

Como ressaltado por Santos et al. (2013), foi observado que existe uma forte influência da disciplina cursada e do docente que a ministra sobre a evasão discente percebida pela mineração de dados. Esta influência acontece não apenas na dimensão didático-pedagógica, facilitando ou dificultando ao aluno a conclusão de uma disciplina, prevista como uma causa da evasão pela Comissão Especial de Estudos sobre Evasão nas IES Públicas (BRASIL/MEC, 1997), mas também na relação entre o professor e o SAE, tornando os dados dos alunos armazenados por este assistente virtual (SAE) mais ou menos ricos.

Foi possível inferir que a postura dos docentes C, D, E e F se mantém mais tradicionais, ou seja, trabalham o processo de ensino centrado no professor, enquanto que os docentes A e B labutam a maior autonomia em seus aprendizes e trabalham o processo educacional mais centrado na aprendizagem de cada estudante. Um tópico interessante para estudos futuros é o aprofundamento na análise da postura docente que usa este tipo de software educacional (SAE), capaz de trabalhar a aprendizagem ativa e personalizada de cada aprendiz, de maneira muito tradicional (ensino centrado no professor), possivelmente, os desestimulando e incentivando a certa dependência do professor, podendo este também ser um fator importante à evasão no ensino superior.

## 6. REFERÊNCIAS

ALI, A., SHUBRA, C., "Efforts to Reverse the Trend of Enrollment Decline in Computer Science Programs", *Issues in Informing Science and Information Technology*, Vol. 7, 2010, pp. 209-225.

CAMPELLO, A. V. C; LINS, L. N. Metodologia de análise e tratamento da evasão e retenção em cursos de graduação instituições federais de ensino superior. In: *Anais XXVIII Encontro Nacional de Engenharia de Produção*, Rio de Janeiro, 2008.

COBBE, P. R; BALANIUK, R; PRADO, H. A; GUADAGNIN, R. V; FERNEDA, E. Predicting evasion candidates in higher education institutions. *Model and Data Engineering*, Brasília, v. 6918, p. 143-151, 2011.

COMISSÃO ESPECIAL DE ESTUDOS SOBRE EVASÃO NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS. BRASIL. MEC. *Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas*. 1997.

DEKKER, G, PECHENIZKIY, M; VLEESHOUWERS, J. Predicting Students Drop Out: A Case Study. In *Anais Proceedings of the International Conference on Educational Data Mining*, Córdoba, Espanha, 2009, p. 41-50.

DORE, R; LÜSCHER, A. Z. Permanência e evasão na educação técnica de nível médio em Minas Gerais. *Cad. Pesqui.*, São Paulo, v. 41, n. 144, 2011.



GAIOSO, N. P. L. Evasão discente na educação superior: a perspectiva dos dirigentes e dos alunos. Brasília: Programa de Mestrado em Educação, 2005. 99 p. Dissertação (Mestrado) Universidade Católica de Brasília, 2005.

MARTINS, C. B. N. Evasão de alunos nos cursos de graduação em uma instituição de ensino superior. Pedro Leopoldo: Programa de Mestrado Profissional de Administração, 2007. 116p. Dissertação (Mestrado Profissional) Fundação Dr. Pedro Leopoldo, 2007.

RISSOLI, V. R. V; SANTOS, G. A. Um Assistente Inteligente Fuzzy no Acompanhamento da Aprendizagem Significativa. In: Congresso da Sociedade Brasileira de Computação, 2011, Natal. Anais do Congresso da Sociedade Brasileira de Computação.

RISSOLI, V. R. V. Uma proposta metodológica de acompanhamento personalizado para Aprendizagem Significativa apoiada por um Assistente Virtual de Ensino Inteligente. Porto Alegre: PGIE, 2007. 224 p. Tese (Doutorado) Universidade Federal do Rio Grande do Sul, 2007.

SANTOS, A. P; SANTOS, S. I. F; RISSOLI, V. R. V. A Predição da Evasão de Estudantes de Graduação como Recurso de Apoio Fornecido por um Assistente Inteligente. In: Simpósio de Excelência de Gestão e Tecnologia, 2013, Resende. Anais do Simpósio de Excelência de Gestão e Tecnologia.

SILVA, H. M; MARCIANO, L. A; BATISTA, I. D. S; GORGONIO, F. L. Uma Reflexão Sobre o Crescente Desinteresse e a Constante Evasão em Cursos de Computação e Informática. In: Intertech'2012 - XII International Conference on Engineering and Technology Education, 2012, Timor Leste. Engineering and Technology Education: Turning Challenges into Opportunities. São Paulo: COPEC, 2012. v. 1. p. 39-39.

SUPERBY, J; VANDAMME, J.-P; MESKENS, N. Determination of factors influencing the achievement of the first-year university students using data mining methods. In Proc. of the Workshop on Educational Data Mining at ITS'06, p. 37-44, 2006.

WASHIO, T; SHINNOU, Y; YADA, K; MOTODA, H; OKADA, T. Analysis on a Relation Between Enterprise Profit and Financial State by Using Data Mining Techniques. Springer v. 4384, p.306-316, 2007.

WITTEN, I. H; FRANK, E. Data mining: practical machine learning tools and techniques. 2 ed. San Francisco: Elsevier, 2005.

YACEF, K. Intelligent Teaching Assistant Systems. In: International Conference on Computers in Education, 2002. New Zeland. Anais de Proceedings International Conference on Computers in Education. New Zeland: IEEE, 2002. v. 1, p. 136-140.