



IDENTIFICAÇÃO DE PADRÕES DE EVASÃO DE CURSOS DE ENGENHARIA A PARTIR DE DADOS DE DESEMPENHO ACADÊMICO

Sandro da Silva Camargo – sandro.camargo@unipampa.edu.br

Juliano Marcuzzo Borin – juliano.borin@gmail.com

Érico Marcelo Hoff do Amaral – erico.amaral@unipampa.edu.br

Ana Paula Lüdtke Ferreira – ana.ferreira@unipampa.edu.br

Universidade Federal do Pampa (UNIPAMPA)

Travessa 45, número 1650

CEP 96413-170– Bagé – RS

Resumo: *A evasão é um problema que está presente na educação superior em todo o mundo. Este problema gera perdas não só acadêmicas mas também sociais e econômicas. A obtenção de dados mais exatos, consistentes, oportunos e contextualizados pode ajudar a identificar padrões entre estudantes que estão evadindo das diferentes instituições e as razões que contribuem para estas decisões críticas em suas vidas. A compreensão deste fenômeno é indispensável para a criação de ações para evitá-lo e conduzir a um uso apropriado dos recursos educacionais das instituições de ensino. Entretanto, devido à grande quantidade de dados relativos a cada estudante em diferentes sistemas de informação das universidades, alidada a grande quantidade de estudantes, a análise manual destes dados é proibitiva. Dentro deste escopo, o presente trabalho descreve o processo de identificação de padrões entre alunos que evadiram, a partir dos dados históricos de seu desempenho acadêmico. Este processo englobou o desenvolvimento de um software de transformação e integração de dados dos sistemas de informação institucionais, seguido pelo uso de uma ferramenta de mineração de dados e pela análise dos padrões descobertos. A fim de validar todo o processo, foi executado um estudo de caso usando dados de estudantes de cursos de Engenharia de um dos Campi da Universidade Federal do Pampa. Foram encontrados diversos padrões que permitem predizer quais alunos irão evadir, com taxa de acerto de até 93%, nos diferentes cursos de Engenharia de um Campi da instituição.*

Palavras-chave: *Padrões de evasão, Cursos de Engenharia, Mineração de Dados*

1. INTRODUÇÃO

A evasão é um dos problemas mais severos da educação superior ao redor do mundo. No Brasil, estimativas apontam que a evasão do ensino superior gerou, em 2009, prejuízos financeiros em torno de R\$ 9 bilhões (NOGUEIRA, 2014). Especificamente nas universidades federais, pesquisas apontam que cerca de 13,2% dos alunos evadiram em 2010 (BORGES, 2014). Tais dados geram grande preocupação para as instituições de ensino



superior, onde causam perdas irreparáveis. Além disso, a evasão também tem impacto negativo no desenvolvimento do país, pois as vagas deixadas por estudantes evadidos, na maioria das vezes, não serão ocupadas novamente. Como consequência, a evasão diminui a quantidade de profissionais formados anualmente no Brasil (SILVA FILHO *et al.*, 2007). Tendo em vista a carência de profissionais qualificados no mercado nacional, especialmente na área de Engenharia, são potencializadas as consequências negativas geradas pela evasão (Falta, 2014).

O prejuízo também pode ser analisado pelo ponto de vista econômico, onde existem receitas, definidas de acordo com a quantidade de alunos, destinadas para manutenção da estrutura física, contratação de professores e profissionais ligados às universidades, e que não são aproveitadas em sua totalidade (SILVA FILHO *et al.*, 2007). Neste contexto, um melhor aproveitamento dos recursos dedicados ao ensino superior está diretamente ligado ao combate à evasão. Assim, entender e estudar quais fatores estão ligados a evasão é de vital importância para uma melhora da educação no Brasil.

São muitos os fatores que podem levar alunos a desistirem de seus estudos. Existem fatores desde dificuldades financeiras, de adaptação à uma nova cidade ou a escolha de uma carreira que não atende as expectativas criadas previamente (ALVES & ALVES, 2010). Essas situações não estão ligadas diretamente à vida acadêmica do aluno, sendo mais difícil identificar padrões que envolvam-nas nos perfis dos evadidos, justamente pela ausência de coleta e armazenamento destes dados nos sistemas de informação institucionais. Também existem fatores educacionais, como dificuldades de aprendizagem em determinados assuntos, professores com deficiências didáticas e reprovações (SILVA FILHO *et al.*, 2007).

Este trabalho defende que os dados armazenados nos sistemas de informação das Instituições de Ensino Superior (IES) podem ser utilizados para a identificação de perfis de evasão e, a partir deles, a previsão de quais alunos estão propensos a abandonarem seus estudos. Devido à quantidade de alunos evadidos nas instituições e à quantidade de dados sobre estes alunos, armazenados nos diversos sistemas de informação das universidades, a utilização de um método manual para a identificação e análise dos perfis dos alunos evadidos é impraticável. Dessa maneira, torna-se imprescindível a utilização de técnicas e ferramentas que permitam a automatização do processo, possibilitando rapidamente a identificação dos padrões de evasão e a tomada das decisões necessárias. Dentro deste escopo, este trabalho visa apresentar um estudo de caso sobre um processo de identificação de padrões de evasão a partir de dados de desempenho acadêmico dos alunos. O contexto da execução da proposta é uma instituição pública de ensino superior brasileira, com foco em cursos de Engenharia.

O restante do trabalho está organizado da seguinte forma: a Seção 2 descreve o contexto onde o trabalho foi aplicado; a Seção 3 aborda a metodologia para o desenvolvimento do trabalho; a Seção 4 apresenta e discute os modelos descritivos de padrões de evasão identificados no estudo de caso; a Seção 5 apresenta uma visão geral dos resultados obtidos e aponta os trabalhos futuros.

2. CONTEXTO

Nas últimas duas décadas o Brasil tem vivenciado uma expressiva elevação no número de vagas oferecidas no ensino superior. Um dos fatores que contribuiu para esta mudança foi o programa de expansão da Rede Federal de Educação Superior, que teve início em 2003, com a interiorização dos *campi* das universidades federais. Dentro deste contexto, em 2006, a



Universidade Federal do Pampa foi criada pelo Governo Federal, com o objetivo de minimizar o processo de estagnação econômica da metade sul do Rio Grande do Sul. Este objetivo está baseado na crença de que a educação viabiliza o desenvolvimento regional. Desta forma, esta nova universidade busca ser um agente da definitiva incorporação da região ao mapa do desenvolvimento do estado. A Unipampa é uma universidade multicampi, contando com 10 *campi* distribuídos em diferentes cidades da metade sul e fronteira oeste do Rio Grande do Sul. Como um primeiro reconhecimento do trabalho desenvolvido na instituição, em 2011 a Unipampa recebeu a primeira avaliação do Índice Geral de Cursos (IGC), somando as avaliações presenciais e as notas obtidas no Exame Nacional de Desempenho de Estudantes (ENADE), e obtendo a nota 4 (em um ranking de 1 a 5), ficando na quarta melhor colocação entre as universidades públicas do Rio Grande do Sul. O bom resultado se manteve em 2013, quando manteve a nota 4 no IGC.

Até o ano de 2013, a Unipampa contabilizava quase 21.000 ingressantes ao longo de sua curta história de 8 anos, tendo formado aproximadamente 2.500 alunos. Porém, apesar de ter potencialmente em torno de 18.500 alunos matriculados, o número efetivo de matrículas era pouco superior a 9.000. Estes números demonstram uma realidade extremamente preocupante, dados os mais de 9.500 casos de evasão na história da universidade. A Figura 1 mostra a evolução destes números desde a criação da Unipampa até o ano de 2013. Pode ser observado que a quantidade de alunos regulares tem apresentado um crescimento menor do que o apresentado pela quantidade de ingressantes. Em 2013, a Unipampa ofereceu mais de 3.700 vagas, distribuídas entre os 63 cursos de graduação da instituição. Porém, descontadas as vagas desocupadas devido à evasão, o crescimento efetivo foi de apenas 1.400 alunos regulares. O problema da evasão também é sentido de forma generalizada no ensino superior brasileiro (NOGUEIRA, 2014), (BORGES, 2014), (SILVA FILHO *et al.*, 2007), (Falta, 2014), mostrando que se a política de expansão do ensino superior se baseia no incremento de vagas nas instituições, a identificação dos fatores que influem na permanência e na evasão do sistema de ensino é fundamental para a otimização desta política.

A partir da identificação do problema de alta evasão na universidade, o primeiro passo realizado foi a análise dos dados de alunos formados e evadidos nos seus diferentes *campi*. Para otimização de recursos, cada um dos 10 *campi* da Unipampa é dedicado a uma ou mais áreas de conhecimento, conforme mostrado na Tabela 1. Esta tabela também apresenta a quantidade total de alunos que já ingressaram em cada um dos *Campus*, assim como a quantidade de alunos formados e evadidos desde a criação da Universidade. Os percentuais de alunos formados e evadidos permitem uma análise relativa destes indicadores, sendo enfatizados os dados das duas unidades com menor percentual de alunos formados e das duas com maior percentual de alunos evadidos. Pode ser verificado que os *campi* que menos formam alunos são os que oferecem cursos nas áreas das Engenharias. Como agravante, o Campus Bagé também possui o com maior percentual de evasão na Universidade. Por este motivo, este Campus foi escolhido como foco do estudo apresentado neste trabalho.

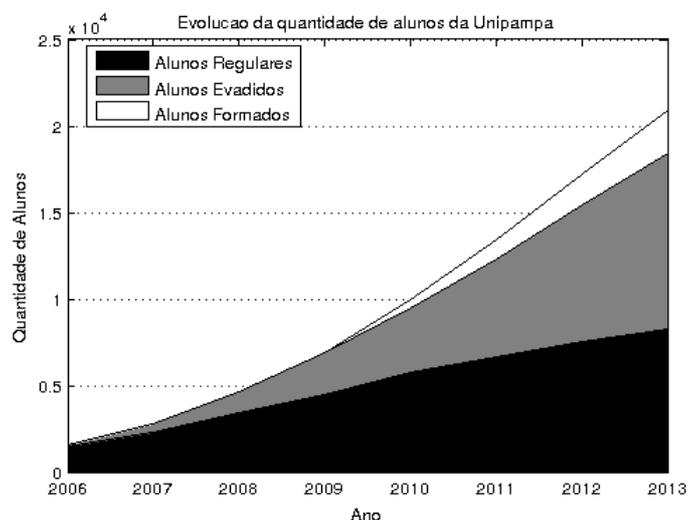


Figura 1 – Evolução da quantidade de alunos da Unipampa

Tabela 1 – Quantidade de alunos formados e evadidos por Campus/Área

Campus	Ingressantes	Evadidos	Formados	%		Área
				Evadidos	Formados	
Alegrete	2254	1031	150	45.7	6.7	Engenharias
Bagé	4633	2829	325	61.0	7.0	Engenharias e Licenciaturas
Caçapava do Sul	998	425	75	42.5	7.5	Exatas e da Terra
Dom Pedrito	931	365	111	39.2	11.9	Agrárias
Itaqui	1696	745	137	43.9	8.1	Agrárias
Jaguarão	1657	813	268	49.0	16.2	Humanas
Santana do Livramento	2142	1012	288	47.2	13.4	Sociais Aplicadas
São Borja	1923	799	428	37.3	22.3	Sociais Aplicadas
São Gabriel	1708	916	266	53.6	15.6	Biológicas
Uruguaiana	2650	941	410	35.5	15.5	Saúde
Total	20592	9876	2458	47.9	11.9	

Uma observação mais detalhada sobre as formas que os alunos deixam a Unipampa pode ser vista na Figura 2. É possível verificar a importância relativa das formas mais severas de evasão. A grande representatividade dos abandonos neste gráfico mostra que este deve ser o principal foco de concentração de esforços para a minimização dos índices de evasão na Unipampa.

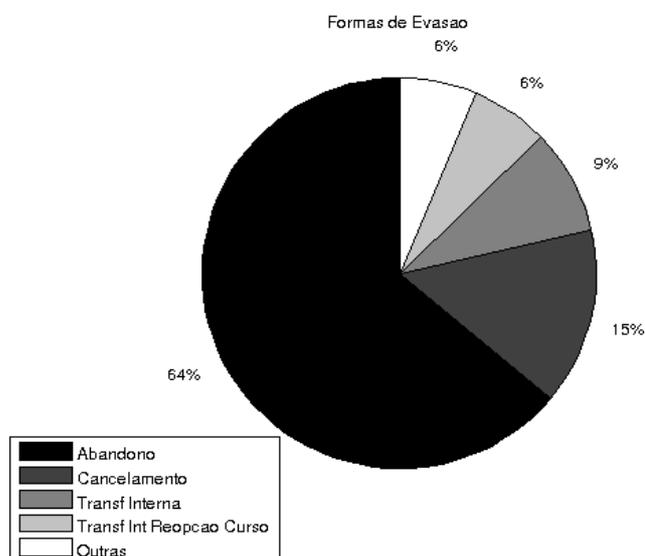


Figura 2 – Formas de evasão dos alunos da Unipampa

Conforme comentado anteriormente, a análise sobre os níveis de evasão e de alunos formados dos diferentes *campi*, apresentados na Tabela 1, motivou a escolha do Campus Bagé para a realização deste trabalho. Este Campus oferece atualmente 11 cursos, sendo 6 Licenciaturas e 5 Engenharias. Os cursos de Engenharia oferecidos são: Engenharia de Alimentos, de Computação, de Energias Renováveis e Ambiente, de Produção e Engenharia Química. Os dados apresentados na Tabela 1 são detalhados na Tabela 2 para os cursos de Engenharia deste Campus. Os dados enfatizados nesta tabela indicam o percentual de formados menor que a média do Campus, assim como o percentual de evadidos maior que a média do Campus.

Tabela 2 – Alunos formados e evadidos nos cursos de Engenharia do Campus Bagé

Curso	Ingressantes	Formados	Evadidos	% Formados	% Evadidos
Engenharia de Alimentos	377	21	258	5.5	68.4
Engenharia de Computação	474	18	288	3.7	60.7
Engenharia de Energias Renováveis e Ambiente	424	25	262	5.8	61.7
Engenharia de Produção	556	35	283	6.2	50.8
Engenharia Química	445	41	192	9.2	43.1

3. METODOLOGIA

A fim de permitir a investigação dos fatores acadêmicos, que contribuem para a evasão de alunos na Unipampa, este trabalho propõe a utilização de técnicas de mineração de



dados, aplicadas sobre os dados históricos de desempenho acadêmico dos alunos. Para realizar este processo, foram realizadas as atividades previstas no padrão *Cross Industry Standard Process for Data Mining* (CRISP-DM) (SHEARER, 2000).

Na fase de compreensão do problema, foi definida como finalidade principal do processo a identificação de padrões dentre os dados dos alunos evadidos da Unipampa. Trata-se, pois, de um problema de classificação, onde o foco dos algoritmos de mineração de dados consiste em classificar a cada final de semestre, os alunos propensos a evadirem. Na fase de preparação dos dados envolveu a criação de um software para integrar os dados das diferentes fontes institucionais, SIE (Sistema de Informação, 2014) e SISU (Sistema de Seleção, 2014), e a transformação destes dados para o formato ARFF (Weka, 2014), requerido pela ferramenta Weka para aplicação dos diferentes algoritmos de mineração de dados. O software desenvolvido para realizar a preparação, descrito em BORIN (2014), foi chamado de MineraPampa e visa realizar as atividades de limpeza, integração, redução e transformação dos dados originais em um formato potencialmente aproveitável pelos diferentes algoritmos de mineração que serão aplicados na etapa seguinte. Isto é possível porque o MineraPampa permite a configuração de parâmetros para preenchimento de valores inexistentes com valores "0" ou "?", preenchimento dos resultados das disciplinas com nota ou com aproveitamento ("Aprovado", "Reprovado por Nota", "Reprovado por Frequência"), e contador de vezes que a disciplina foi cursada. Na fase de construção dos modelos, foi utilizada a ferramenta Weka (Weka, 2014), que implementa uma coleção de algoritmos de mineração de dados. Foi priorizada a construção de modelos do tipo caixa branca, onde se enquadram os algoritmos de geração de regras ou árvores de decisão. Apesar destes modelos serem lineares e, tipicamente, terem uma capacidade preditiva menor que modelos não lineares, sua maior vantagem é a geração de modelos interpretáveis (CAMARGO, 2010). Na fase de avaliação dos modelos foi utilizada a técnica *10-fold cross validation* para assegurar a correta mensuração da capacidade preditiva dos modelos. Em resumo, os algoritmos de mineração de dados são aplicados sobre os dados históricos de desempenho acadêmico dos alunos, desde seu ingresso na Universidade, seus dados de desempenho no ENEM, e de outros dados obtidos a partir do cadastro do aluno.

A fase seguinte, será a utilização dos modelos, onde se pretende monitorar os alunos que se enquadram dentro dos padrões de evasão identificados na fase anterior, a fim de tentar prevenir que abandonem seus cursos. Também está sendo discutida a criação de políticas institucionais para prevenção da evasão por um Grupo de Trabalho específico para este assunto.

4. PADRÕES DE EVASÃO DESCOBERTOS E DISCUSSÃO DOS RESULTADOS

Após a preparação dos dados foi iniciada a etapa de identificação dos padrões de evasão, onde são utilizados os algoritmos de mineração sobre os dados. Esta etapa foi executada com a ferramenta Weka, com ênfase principal na utilização de algoritmos de classificação, na tentativa de identificar padrões entre os dados dos estudantes que evadiram da universidade. Os experimentos foram realizados utilizando as configurações possíveis dos parâmetros do MineraPampa para aumentar as possibilidades de uso dos diferentes algoritmos disponíveis no Weka. Foram realizados experimentos para identificação de padrões de diferentes cursos de Engenharia do Campus Bagé da Unipampa.

Em relação ao curso de Engenharia de Alimentos, foram identificados 5 padrões de evasão diferentes. Um dos padrões encontrados, mostrado na Figura 3a, engloba uma disciplina básica do curso. Apesar da regra explicar apenas 12 casos de evasão do curso, ela tem uma taxa de acerto de 100%.

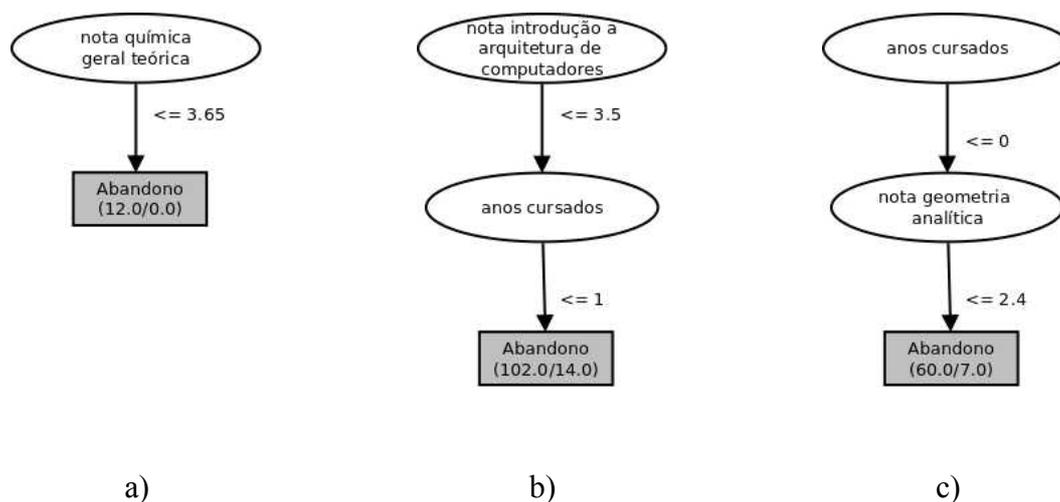


Figura 3 – Padrões de evasão dos cursos de a) Engenharia de Alimentos, b) e c) Engenharia de Computação

Em relação ao curso de Engenharia de Computação, foram identificados 9 padrões de evasão. Um dos padrões, mostrado na Figura 3b, apresenta uma relação entre a nota na disciplina de Introdução a Arquitetura de Computadores, que é um dos alicerces do curso, e o abandono. Entre os 102 alunos que obtiveram notas iguais ou inferiores a 3.5 nesta disciplina e cursaram até um ano, ou 2 semestres no curso, apenas 14 de 102 não abandonaram os estudos, representando mais de 86% de taxa de acerto desta regra. O padrão apresentado na Figura 3c descreve uma relação entre abandono e reprovação na disciplina de Geometria Analítica, que é ofertada no primeiro semestre do curso. Dentre os 60 alunos que ainda concluíram somente o primeiro semestre do seu curso e obtiveram nota igual ou inferior a 2.4 nesta disciplina, apenas 7 não abandonaram a universidade, indicando uma taxa de acerto desta regra de mais de 88%.

Em relação ao curso de Engenharia de Energias Renováveis, foram encontrados 9 padrões de evasão. O padrão mostrado na Figura 4a mostra alunos que já cursaram a disciplina de Química Geral Experimental, mas ainda não foram aprovados, e ingressam no curso antes do ano de 2010 incorreram em abandono. Dentre os 65 alunos que se enquadraram nesta regra, 58 evadiram, o que indica uma taxa de acerto superior a 89% para esta regra.

No tocante ao curso de Engenharia de Produção, foram encontrados 7 padrões de evasão diferentes. A Figura 4b mostra uma das regras encontradas. Esta regra indica que alunos cursaram até três anos e obtiveram notas iguais ou inferiores a 6.3 em Gestão da Qualidade II e 6.1 em Calculo II têm uma forte tendência de abandonarem o curso. Esta regra explica 105 dos 121 abandonos que se enquadram nela, apresentando 87% de acerto.

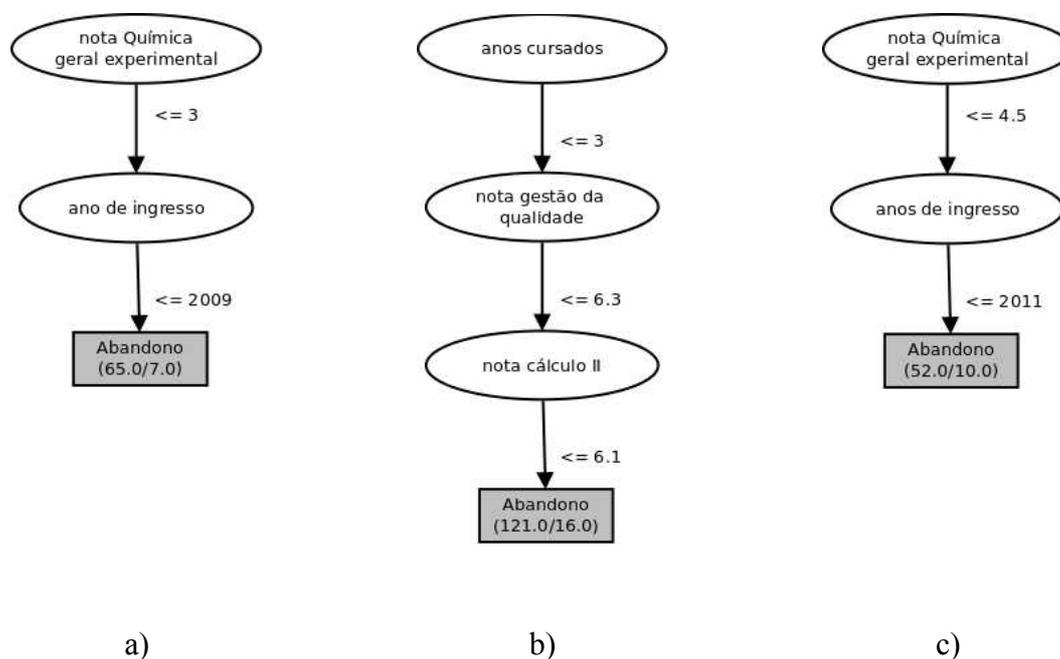


Figura 4 – Padrões de evasão dos cursos de a) Engenharia de Energias Renováveis e Ambiente, b) Engenharia de Produção e c) Engenharia Química

No curso de Engenharia Química, foram identificados 5 padrões de evasão. Na Figura 4c é apresentado um dos padrões encontrados, o qual indica que entre os 52 alunos que ingressaram até 2011 em Engenharia Química e reprovaram em Química Geral Experimental com notas variando entre 0 e 4.5, ocorreram 42 casos de evasão. Tal regra apresenta uma taxa de acerto de 80%.

Os padrões descritos anteriormente representam apenas alguns dos padrões identificados em cada curso. Um resumo de todos os padrões de evasão encontrados é apresentado na Tabela 3. A tabela mostra os algoritmos que foram utilizados, a precisão geral do modelo criado, que indica a capacidade do modelo prever alunos evadidos, formados, regulares, abandonos, desligamentos e transferências; e a estatística de kappa, que permite mensurar a qualidade do modelo. Valores de kappa mais próximos de 1 indicam uma melhor qualidade do modelo criado. Tais valores mostram que os modelos de evasão criados não são apenas fruto do acaso, tendo validade estatística. A tabela também mostra o resultado específico para identificação de alunos que abandonaram seus cursos, nas colunas referentes à classe Abandono. Uma descrição detalhada de todos os padrões identificados pode ser obtida em BORIN (2014). Na coluna cfg é indicada a configuração utilizada pelo MinerPampa durante o processo de transformação dos dados originais para o formato a ser utilizado pela ferramenta Weka. O valor 1 na coluna indica que valores omitidos foram substituídos pelo sinal “?”, o resultado que o aluno obteve na disciplina é descrito através da sua nota, e não foi incluída a quantidade de vezes que o aluno cursou cada disciplina. O valor 2 indica o uso dos parâmetros ?, Notas, quantidade de vezes. O valor 3 indica o uso de ?, Aproveitamento, com quantidade. O valor 4 indica que valores omitidos foram substituídos por 0, e os demais parâmetros foram Notas, e sem o contador. O valor 5 indica os parâmetros 0, Notas, com os contadores.

Tabela 3 – Resumo dos padrões identificados nos cursos de Engenharia do Campus Bagé.

Curso	Modelo				Classe “Abandono”		
	Algoritmo	Cfg	Precisão Geral	Kappa	Casos corretamente classificados	Precisão	Precisão Média
Engenharia de Alimentos	ASC	1	63.95%	0.5132	28 de 31	90.32%	72.25%
	ASC	4	60.46%	0.4887	24 de 31	77.42%	
	FC	1	58.13%	0.4662	20 de 31	64.52%	
	FC	4	59.30%	0.4739	23 de 31	74.19%	
	Ridor	1	53.48%	0.4173	17 de 31	54.84%	
Engenharia de Computação	FC	1	74.67%	0.5693	150 de 175	85.71%	89.07%
	FC	2	77.29%	0.6229	162 de 175	92.57%	
	JRip	1	81.65%	0.6920	158 de 175	90.29%	
	JRip	3	81.22%	0.6887	155 de 175	88.57%	
	JRip	4	83.18%	0.7204	160 de 175	91.43%	
	JRip	5	81.00%	0.6823	152 de 175	86.86%	
	PART	1	84.06%	0.7375	163 de 175	93.14%	
	PART	4	82.09%	0.7148	147 de 175	84.00%	
	J48	2	81.87%	0.7019	156 de 175	89.14%	
Engenharia de Energias Renováveis e Ambiente	FC	1	73.00%	0.5763	109 de 135	80.74%	72.83%
	FC	3	70.00%	0.5256	109 de 135	80.74%	
	JRip	1	71.25%	0.5521	107 de 135	79.26%	
	JRip	3	75.50%	0.6233	112 de 135	82.96%	
	JRip	4	72.75%	0.5773	110 de 135	81.48%	
	JRip	5	71.25%	0.5570	10 de 135	7.41%	
	PART	1	76.75%	0.6469	109 de 135	80.74%	
	PART	3	75.50%	0.6187	111 de 135	82.22%	
	PART	4	74.25%	0.6143	108 de 135	80.00%	
Engenharia de Produção	FC	1	77.78%	0.6189	171 de 195	87.69%	85.56%
	ASC	1	74.81%	0.5679	126 de 195	64.62%	
	JRip	1	83.33%	0.7225	171 de 195	87.69%	
	JRip	4	82.59%	0.7085	169 de 195	86.67%	



	PART	1	86.29%	0.7742	179 de 195	91.79%	
	PART	3	85.18%	0.7547	175 de 195	89.74%	
	J48	4	86.85%	0.7851	177 de 195	90.77%	
Engenharia Química	ASC	4	78.80%	0.6596	68 de 95	71.58%	68.42%
	FC	1	75.47%	0.5958	62 de 95	65.26%	
	JRip	1	72.85%	0.5364	54 de 95	56.84%	
	PART	1	80.47%	0.6862	72 de 95	75.79%	
	PART	4	77.61%	0.6483	69 de 95	72.63%	

Na coluna precisão média está indicado a confiança dos padrões de evasão, obtidos a partir dos históricos dos alunos, para casos futuros, de acordo com cada um dos cursos de Engenharia estudados. Com base nesta coluna é possível inferir que os cursos de Engenharia de Computação e de Produção há padrões de evasão mais bem definidos que nos demais cursos. Nos cursos de Engenharia de Alimentos, de Energias e Química os padrões existentes são menos precisos mas, de qualquer forma, podem fornecer importantes subsídios para a criação de políticas de prevenção de abandonos.

Também pode ser observado que diferentes algoritmos, aplicados sobre os mesmos dados, podem gerar modelos diferentes. Isso ocorre porque cada algoritmo tem um método diferente para selecionar o atributo mais informativo em um determinado escopo. É importante salientar que vários outros dados dos alunos, tais como sexo, idade, raça, idade e renda, foram utilizados como entrada no processo de criação dos modelos. Porém, nenhum destes dados foi considerado como importante pelos algoritmos de mineração para criação dos modelos. No processo de criação dos modelos foi possível verificar que somente os dados de desempenho acadêmico permitiram a identificação de vários padrões de evasão com taxa de acerto de até 93%.

A existência de padrões de evasão que envolvam disciplinas básicas, tais como Geometria Analítica e Química Geral, pode indicar que a principal causa da evasão é a dificuldade originada pela carência de conhecimento prévio, que deveria ter sido obtido no ensino médio. Esta conclusão vai ao encontro do que foi observado em alguns trabalhos relacionados (CAMARGO *et al.*, 2013). Já a existência de padrões de evasão que envolvam disciplinas específicas dos cursos, pode indicar a escolha de um curso que não está de acordo com a aptidão do estudante.

5. CONSIDERAÇÕES FINAIS

Este artigo apresentou o relato da aplicação de uma metodologia de mineração de dados para identificação de padrões de evasão de alunos de uma instituição de ensino superior a partir de sistemas de informação institucionais com dados acadêmicos dos alunos. Um estudo de caso foi realizado a partir dos cursos de Engenharia de um campus da instituição. Para contribuir no processo de mineração, foi desenvolvida uma ferramenta, chamada MinerPampa, que permite que os arquivos gerados a partir do SIE da Unipampa pudessem passar por processos de limpeza, integração, transformação e redução dos dados. A ferramenta também possibilitou a concatenação das notas obtidas pelos estudantes durante a



prova nacional do ENEM e que utilizaram o SiSU como método de entrada na universidade. Porém, as notas do ENEM não foram consideradas, pelos algoritmos utilizados, como atributos informativos para indicar a evasão. Adicionalmente, diferentes possibilidades de parametrização da ferramenta permitiram a geração de arquivos passíveis de utilização por diferentes algoritmos de classificação, possibilitando que uma maior gama de modelos pudesse ser obtida.

Com relação aos experimentos realizados, foi possível criar modelos descritivos dos padrões encontrados a partir dos dados dos alunos evadidos, tornando possível identificar quais fatores relacionados à vida acadêmica dos estudantes contribuem para que ocorra o fenômeno de evasão dentro destes cursos de Engenharia. Foram identificados vários perfis diferentes de evasão, com taxa de acerto de até 93%, que tipicamente indicavam a nota em uma ou mais disciplinas como fator chave para a evasão. Tais regras, aliadas a sua alta precisão, mostram que somente os dados de desempenho acadêmico já permitem a identificação de padrões de evasão altamente representativos. Os diferentes padrões identificados serão utilizados como apoio pelos grupos de trabalho criados dentro da universidade para tentar realizar o combate a evasão. Soluções, de maneira conjunta, entre os diferentes cursos de Engenharia da Unipampa que contribuam para um melhor desempenho dos alunos, como monitorias ou aulas de reforço, em disciplinas que estão presentes em todos os cursos podem ser uma alternativa.

Por fim, a metodologia aqui apresentada pode ser replicada para aplicação em outros cursos de Engenharia de diferentes universidades, a fim de buscar os padrões de evasão em outros contextos.

6. REFERÊNCIAS / CITAÇÕES

ALVES, T. W. e ALVES V. V. Fatores determinantes da evasão universitária: uma análise a partir dos alunos da UNISINOS. Anais: IV Encontro de Economia Catarinense. Cidade: Editora, 2010.

BORGES, P. **MEC e universidades estudam planos para combater evasão.** Disponível em: <<http://ultimosegundo.ig.com.br/educacao/mec-e-universidades-estudam-planos-para-combater-evasao/n1597622390779.html>> Acesso em: 01 jun. 2014.

BORIN, J. M. UNIVERSIDADE FEDERAL DO PAMPA, Campus Bagé. Desenvolvimento de um software para análise de evasão na Unipampa Campus Bagé utilizando técnicas de mineração de dados, 2014. 93 p. il. Monografia (Graduação).

CAMARGO, S. da S.; UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, Instituto de Informática. Um Modelo Neural de Aprimoramento Progressivo para Redução de Dimensionalidade, 2010. Tese (Doutorado).

CAMARGO, S. da S., BETEMPS, C. M. e CECHINEL, C. Contribuições do Projeto IGUAL: Um estudo de caso. Anais: LACLO 2013 – Octava Conferencia Latinoamericana de Objetos y Tecnologías de Aprendizaje. Valdivia: UACH, 2013.



FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v.76, n.5, p. 378-382, 1971.

NOGUEIRA, F. **País perde R\$ 9 bilhões com evasão no ensino superior, diz pesquisador.** Disponível em: <<http://g1.globo.com/educacao/noticia/2011/02/pais-perde-r-9-bilhoes-com-evasao-no-ensino-superior-diz-pesquisador.html>> Acesso em: 01 jun. 2014.

SHARER, C. The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, n.5, p. 13-22, 2000.

SILVA FILHO, R. L. L., *et al.* A Evasão no Ensino Superior Brasileiro. Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia, *Caderno de Pesquisa*, v. 37, n. 132, 2, 2007.

---. **Falta de engenheiros faz com que profissão esteja em alta no Brasil.** Disponível em: <<http://g1.globo.com/jornal-hoje/noticia/2013/03/falta-de-engenheiros-faz-com-que-rofissao-esteja-em-alta-no-brasil.html>> Acesso em: 01 jun. 2014.

---. **Sistema de Informações para o Ensino - SIE.** Disponível em: <<http://www.sie.ufu.br>> Acesso em: 01 jun. 2014.

---. **Sistema de Seleção Unificada - SISU.** Disponível em: <<http://sisu.mec.gov.br>> Acesso em: 01 jun. 2014.

---. **Weka 3: Data Mining Software in Java.** Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/>> Acesso em: 01 jun 2014.

IDENTIFYING DROP OUT PATTERNS IN ENGINEERING COURSES BASED ON ACADEMIC PERFORMANCE DATA

Abstract: *Dropping out is a severe problem present in higher education around the world. This problem generates social, economical and academical losses. There is urgency for obtaining more accurate, consistent, and timely data to identify patterns among students who are dropping out and the reasons contributing to these life-altering decisions. Understanding this phenomena is indispensable to trace effective goals to avoid dropping out and lead to an appropriate use of educational institutions resources. However, due to huge amount of available data regarding each student in different corporative information systems, and the high amount of students, manual analysis of these data is prohibitive. In this scope, this work describes a process aiming identify patterns among students who have dropped out. This process have encompassed the development of a software which have integrated data from different information systems, followed by the use of a data mining tool, called Weka, and the analysis of discovered patterns. In order to validate the process, we have done a study case using data from Engineering students from Universidade Federal do Pampa.*

Key-words: *drop out patterns, engineering courses, data mining*