

# UMA ABORDAGEM PARA PREVISÃO DE EVASÃO EM CURSOS DE GRADUAÇÃO PRESENCIAIS

Gustavo Zanini Kantorski   Ivan Londero Hoffmann   Jader Adiél Schmitt  
Evandro Gomes Flores   Fernando Pires Barbosa  
gustavo@ufsm.br   ilhoffmann@ufsm.br   jaschmitt@cpd.ufsm.br,  
evandro.flores@ufsm.br   fernando.pires.barbosa@gmail.com  
UFSM – Santa Maria – Brasil

**Resumo.** Este artigo apresenta a evasão existente em cursos de graduação de uma Instituição pública de ensino superior. O objetivo é a previsão da evasão em cursos de graduação presenciais com a finalidade de visualizar perspectivas que permitam uma ação efetiva de intervenção, mitigando o processo da evasão. A pesquisa desenvolvida utiliza vários métodos computacionais de previsão, partindo de informações sobre os alunos e seus desempenhos acadêmicos e aplicando técnicas de mineração de dados para a previsão. Os resultados demonstram que a abordagem proposta é factível e eficiente. Os experimentos, em dados reais, alcançaram uma acurácia de 98% na previsão, e mais de 70% de sucesso na previsão de alunos que abandonaram o curso.

**Palavras Chave:** Evasão Escolar, Mineração de Dados, Classificação

**Abstract.** *This paper presents the college dropout that exists on undergraduate programs of public Institutes in Brazil. The aim is the dropout prediction in BSc programs, which helps visualize perspectives of motivation, reducing the dropout process. The developmental research works with several computational prediction methods and contains personal and academic information about the students. The experiments, on real data, demonstrate accuracy around 98% in general prediction, and dropout rate around 70%.*

**Keywords:** *College Dropout, Data Mining, Classification*

## 1. Introdução

A evasão de alunos nos cursos de graduação tem sido um dos problemas relevantes para a gestão acadêmica e financeira das Instituições de Ensino Superior (IES), pois cada vez mais são investidos recursos escassos na atração e captação de alunos e os resultados têm sido influenciados negativamente com taxas elevadas de evasão. A desistência do aluno em um curso superior significa prejuízo para si próprio, ao não se diplomar, para o professor, que não atinge sua meta como educador, para a universidade, pelo não atendimento de sua missão, para a sociedade, pelas perdas sociais/ econômicas; e, para a família, pelo sonho não realizado (SILVA FILHO *et al.*, 2007).

A maioria dos recentes trabalhos que trata a evasão no ensino superior tem como finalidade compreender os fatores que causam a evasão (ADACHI, 2009; AMARAL 2013; PALACIO, 2012; PEREIRA JR, 2012). Entre os vários fatores encontrados nesses trabalhos destacam-se a falta de motivação dos alunos e professores, problemas pessoais e socioeconômicos, insatisfação com o curso/instituição, problemas de aprendizagem associados com metodologias de ensino e processos de avaliação, restrições do mercado de trabalho, desconhecimento prévio do curso sobre os alunos, repetências contínuas em disciplinas do curso, nível de estudo anterior a matrícula, entre outros.

Outros trabalhos (ABU-ODA & EL-HALEES, 2015; AMAYA *et al.*, 2014; AZOUMANA *et al.*, 2014; JADRIĆ *et al.*, 2010; OREA *et al.*, 2005) abordam a evasão no sentido de prever situações de evasão no ensino superior. Geralmente, esses trabalhos utilizam técnicas de descoberta de conhecimento em bases de dados (*knowledge discovery in databases*) ou mineração de dados (*data mining*) para realizar a previsão de resultados.

O presente trabalho apresenta uma abordagem inovadora e tem como foco dois pontos interessantes: o primeiro é a previsão da evasão em cursos de graduação presenciais e o segundo é

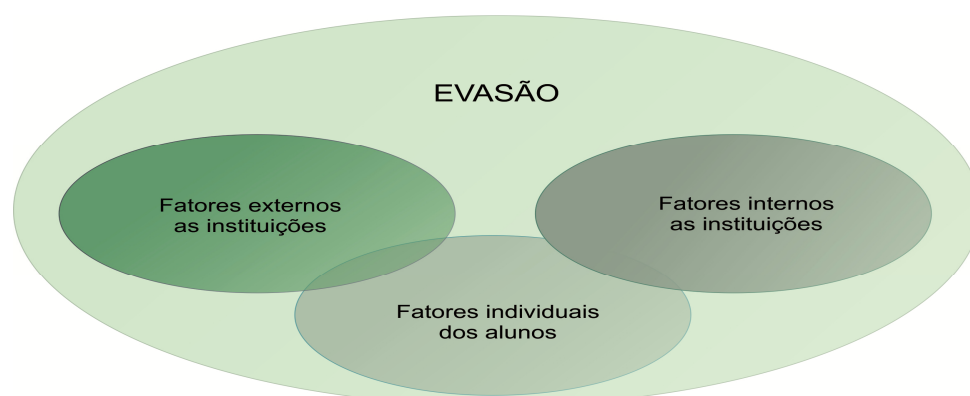
determinar os principais fatores que levam à evasão. O estudo de caso foi realizado para o curso de Zootecnia da Instituição. A intenção é realizar uma análise comportamental dos alunos do curso e determinar, não somente os fatores que contribuem para a evasão, mas também prever os prováveis alunos que possuem tendência à evasão. Nesse sentido, uma análise profunda de atributos de alunos é realizada. Essa análise considera informações pessoais, acadêmicas, de caráter sócio econômico e combinam essas informações para criar um modelo preditivo da evasão. A abordagem proposta nesse trabalho não utiliza somente um modelo de previsão, mas faz uma combinação de modelos de previsão para otimizar o resultado da previsão de possíveis evasões.

## 2. Pontos de Vista da Evasão

A evasão é um dos grandes problemas que cercam o contexto da educação superior, por se tratar de um fenômeno complexo (SCALI, 2009), heterogêneo e macrossocial. Assim, define-se evasão no ensino superior como o movimento do aluno abandonar a IES e nunca receber o diploma (TINTO, 1975), ou mesmo, a interrupção no ciclo dos estudos, em qualquer nível de ensino (MOROSINI *et al.* (2011). Portanto, faz-se necessário compreender o fenômeno da evasão, para poder planejar e criar estratégias para minimizar as perdas que a evasão provoca para todos os envolvidos e em particular para a gestão universitária.

O MEC/SESu (1997), Lobo (2012) por meio da Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras, classifica a evasão no ensino superior em três níveis: evasão de curso, quando o estudante desliga-se do curso de origem sem concluí-lo (abandono, desistência, transferência interna ou aprovação no vestibular para outro curso na mesma instituição); evasão de instituição, quando ele abandona a IES na qual está matriculado (transferência externa ou aprovação no vestibular para curso em outra instituição); e evasão de sistema quando o aluno se ausenta de forma permanente ou temporária da academia (desistência).

O MEC (1996) delimita em três fatores que influenciam a evasão no ensino superior, conforme se apresenta na Figura 1. Os fatores individuais dos alunos referentes à evasão estão relacionados às habilidades de estudo, personalidade, formação escolar anterior, escolha precoce da profissão, dificuldades pessoais de adaptação à vida universitária, desencanto com o curso escolhido, dificuldades recorrentes de reprovações ou baixa frequência e desinformação a respeito da natureza dos cursos. Já os fatores internos às instituições referentes à evasão, podem se caracterizar por questões peculiares à própria academia, a falta de clareza sobre o projeto pedagógico do curso, baixo nível didático-pedagógica, cultura institucional de desvalorização da docência e estrutura insuficiente de apoio ao ensino. Por fim, os fatores externos às instituições como o mercado de trabalho, reconhecimento social na carreira escolhida, conjuntura econômica, desvalorização da profissão, dificuldade de atualizar-se perante as evoluções tecnológicas, econômicas e sociais da contemporaneidade e políticas governamentais. Em outras palavras, evasão escolar trata-se de um fenômeno social complexo definido como interrupção no ciclo de estudos (GAIOSO, 2005), podendo se dar no âmbito do curso, instituição ou sistema.



**Figura 1. Fatores que influenciam a evasão**

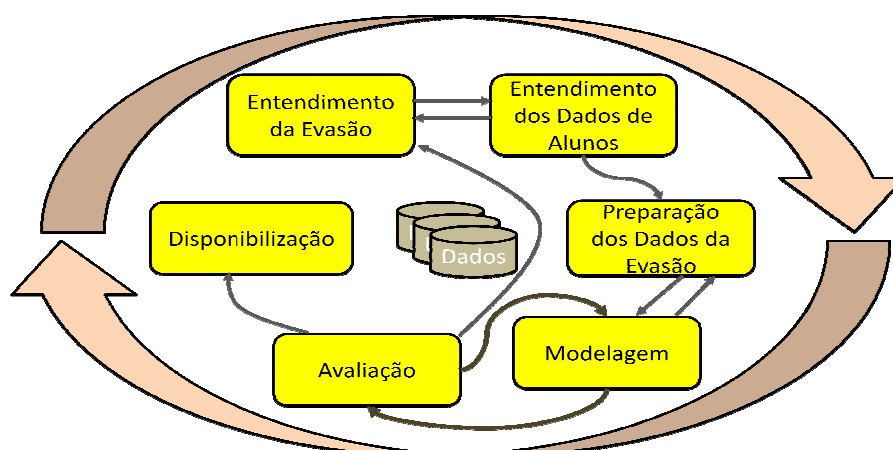
Fonte: Adaptado de MEC (1996)

### 3. Metodologia

A metodologia proposta utiliza conceitos de aprendizagem de máquina para prever as informações relacionadas à evasão. A aprendizagem de máquina é uma área de Inteligência Artificial (IA) cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática (MITCHELL, 1997). Essa área é concentrada na questão de construção de programas de computação que automaticamente melhoram com a experiência. Nos últimos anos muitas aplicações de aprendizagem de máquina foram desenvolvidas com sucesso, tais como, aplicações de mineração de dados para detecção de fraudes em cartões de crédito (CHAN, 1998; CHAN *et al.*, 1999; BRAUSE *et al.*, 1999), sistemas de informação que aprendem as preferências do usuário (AGICHTEIN *et al.*, 2006), veículos que aprendem a dirigir em via pública (POMERLEAU & JOCHEM, 1996), recuperação de informação na Web Oculta/Web Profunda (KANTORSKI & HEUSER, 2012; KANTORSKI *et al.*, 2013), dados associados à evasão escolar (TONTINI, 2013), entre outras. Ao mesmo tempo, ainda existem vários avanços que podem ser realizados em áreas do conhecimento ainda não exploradas.

Dentre as diversas aplicações de sistemas baseados em aprendizagem, uma aplicação especial é a descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases - KDD*). KDD é a extração não trivial de informação implícita, previamente desconhecida e potencialmente útil de um conjunto de dados (FAYYAD *et al.*, 1996). Em outras palavras, é a exploração e análise, de modo automático ou semiautomático, de grandes volumes de dados para descobrir padrões úteis (KANTARDZIC, 2011). Partindo do domínio de aplicação, o especialista avalia os dados particulares do domínio e são gerados dados históricos com a rotulação confirmada pelo especialista. Os dados históricos são organizados por classe, que associam as instâncias do domínio à classe. Por meio da utilização desses dados históricos é construído o conjunto de treinamento. O conjunto de treinamento possui as características do domínio e a respectiva classe. Partindo do conjunto de treinamento um modelo é construído para previsão de novas instâncias do domínio. Assim, quando uma nova instância é avaliada, o especialista não atua no processo. A nova instância é submetida ao modelo que realiza a previsão da classe associada ao domínio.

Para análise da evasão no curso de Zootecnia a metodologia proposta é uma adaptação da metodologia CRISP-DM (*CROSS Industry Standard Process for Data Mining*) (CHAPMAN *et al.*, 2000). A metodologia é geral e não se restringe a uma ferramenta ou tecnologia específica. A Figura 2 ilustra as etapas da metodologia.



**Figura 2. Metodologia CRISP-DM adaptada para Evasão.**

Fonte: adaptada de Chapman (2000).

A primeira etapa é o entendimento da evasão e tem como finalidade identificar os objetivos sob o ponto de vista de descoberta de conhecimento em bases de dados. Esses objetivos incluem o tipo de tarefa a ser executada, por exemplo, previsão e os critérios de avaliação dos modelos

utilizados. Assim, a tarefa a ser executada é a previsão da evasão de alunos do curso de Zootecnia, baseada na análise dos atributos pessoais, acadêmicos, sociais e econômicos dos alunos. Um resultado paralelo diz respeito aos fatores que mais influenciaram na previsão.

A próxima etapa, entendimento dos dados dos alunos, consiste na coleta inicial dos dados e na descrição e exploração dos dados verificando suas propriedades e qualidade. O processo exploratório compreende a análise dos dados e a utilização de técnicas de visualização. A descrição envolve a avaliação de características, tais como, faixas de valores, número de atributos, significado de cada atributo e sua importância para alcançar o objetivo.

A preparação dos dados é a terceira etapa da metodologia. Nessa etapa um conjunto de dados é gerado em conformidade com os modelos de descoberta de conhecimento que serão utilizados. A geração desse conjunto envolve uma série de passos incluindo a seleção, limpeza, construção, integração e formatação dos dados. A seleção dos dados e das instâncias pode ser realizada de maneira manual ou por meio de algoritmos. A limpeza tem como finalidade melhorar a qualidade dos dados. A construção dos dados concentra-se na normalização, se for o caso, na transformação de valores simbólicos para numéricos e na discretização dos atributos.

A etapa de modelagem é a fase de descoberta de conhecimento propriamente dita. Nessa etapa são selecionados os modelos que serão aplicados para a solução da tarefa. A etapa de avaliação interpreta e avalia os resultados em relação aos objetivos de descoberta de conhecimento definidos na etapa de entendimento do negócio. Um ciclo existente entre a etapa de modelagem e avaliação possibilita a avaliação e comparação entre diversos modelos de aprendizagem. Finalmente, a etapa de disponibilização entrega os resultados alcançados por meio da comparação entre os diversos modelos gerados.

#### **4. Aplicação da Metodologia para a Evasão no Curso de Zootecnia**

A metodologia proposta descrita na seção 3 foi aplicada a um curso de graduação presencial da Instituição. Entre os 108 cursos de graduação presencial existentes, o curso de Zootecnia foi selecionado para avaliação. O curso de Zootecnia foi selecionado baseado na avaliação de indicadores do Ministério da Educação (MEC) do Brasil. Uma das variáveis utilizadas para a escolha do curso foi o peso. O peso determina uma área de custo do curso para que seja possível diferenciar cursos de maiores custos em relação aos demais. Por exemplo, um dos fatores predominantes é a utilização de laboratórios que exigem maiores recursos para seu funcionamento. O curso de Zootecnia, juntamente com os cursos de Medicina, Medicina Veterinária e Odontologia, é um dos cursos que possui maior peso. Na perspectiva de evasão na Instituição, nota-se que o maior percentual de evasão entre os cursos com maior peso, é o curso de Zootecnia.

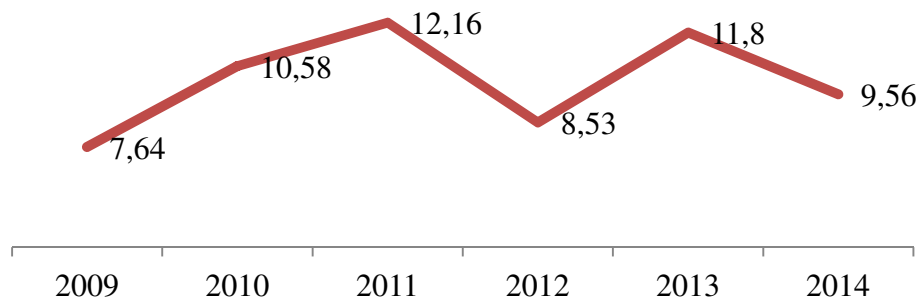
Para o cálculo da evasão foi utilizada a fórmula do Leal (2007) baseado numa comparação entre o número de alunos que estavam matriculados num curso num determinado ano, subtraídos os concluintes, com a quantidade de alunos matriculados no ano seguinte, subtraindo deste último o total de ingressantes desse ano, conforme a Equação  $E_n = 1 - \frac{[M_n - I_n]}{[M_{n-1} - C_{n-1}]}$ , em que:

- |                              |                              |                                       |
|------------------------------|------------------------------|---------------------------------------|
| • E = taxa de evasão anual   | • I = número de ingressantes | • C = número de concluintes           |
| • M = número de matriculados | • n = ano em estudo          | • (n-1) = ano anterior ao considerado |

Este modelo avalia a evasão no período de um ano e os insumos para o seu cálculo envolve o ano que se quer calcular e o ano seguinte. A Figura 3 apresenta a taxa de evasão do curso de Zootecnia, desde 2009 até 2014 e o valor médio da taxa de evasão é 10,04, nos últimos seis anos. O gráfico leva em consideração todas as formas de ingresso, por exemplo, vestibular, transferência, portador de diploma, reingresso, entre outras.

A metodologia proposta foi aplicada em três simulações. Uma simulação possui um conjunto de atributos e um conjunto de modelos. O que diferencia uma simulação de outra são os atributos selecionados para o processo preditivo e os modelos de descoberta de conhecimento aplicados a esses atributos. A primeira simulação consiste de 25 atributos, a segunda simulação

contém 32 atributos e a terceira simulação é composta de 34 atributos. Além dos atributos selecionados, existe um atributo especial que é o atributo classe. O atributo classe possui dois valores para avaliação: Aluno Regular e Evadido. Além disso, em uma mesma simulação, foram utilizadas técnicas para normalização dos atributos. A normalização pode ser necessária, pois alguns atributos assumem uma faixa de valores mais ampla do que outros e, determinados modelos podem ser influenciados por atributos com valores mais altos ou mais baixos. A normalização faz com que todos os atributos fiquem na mesma faixa de valores.



**Figura 3. Taxa de evasão – Curso Zootecnia**

Fonte: SIE – Sistemas de Informação da Instituição (2015)

Considerando as etapas de entendimento e preparação dos dados da metodologia proposta, a coleta de dados levou em consideração vários tipos de atributos, entre eles, atributos pessoais, acadêmicos, sociais, econômicos e financeiros. Essas categorias foram selecionadas conforme vários trabalhos (TIGRINHO, 2008; SILVA FILHO, 2007; SOBRINHO, 2008), existentes na literatura sobre o tema. O trabalho de (LOBO, 2012) aborda as causas mais comuns existentes na evasão. Os atributos selecionados para aplicação da metodologia proposta são baseados nas causas gerais e em questões levantados por (TINTO, 1982), as quais envolvem o acompanhamento do aluno junto ao curso. Cabe salientar que uma limitação na definição dos atributos está relacionada na disponibilidade de outras informações confiáveis nos sistemas de informação projetados para atender as necessidades de controle das atividades desenvolvidas pelo acadêmico na Instituição. Exemplos de atributos que, geralmente, não são registrados, dizem respeito aos aspectos sociais e emocionais.

As informações de alunos com ingresso no ensino superior de 2008 a 2014 foram consideradas para análise. No período indicado foram selecionados 394 instâncias de treinamento. Do ponto de vista da categoria pessoal, do total de instâncias de treinamento, 217 são do sexo feminino e 177 do sexo masculino. Ainda, 292 instâncias não ingressaram por sistema de cotas, enquanto que 102 cotistas estão no conjunto. A análise sócio econômica mostra que a maioria dos alunos é proveniente de escola pública (252), assim como a maioria dos alunos irá se manter com recursos familiares durante o curso (273). A maioria das informações da categoria Institucional é faltante (em torno de 90%). Isso ocorre porque essas informações são geradas no processo de avaliação institucional, em que a maioria dos alunos não participa ativamente. Um fato observado é que 90% das instâncias previstas não responderam ao processo de avaliação institucional.

O viés de assistência estudantil descreve que existem mais alunos bolsistas (56%), que não bolsistas. Na perspectiva de bolsas de monitoria a situação é inversa. A maioria dos alunos não possui bolsa de monitoria (95%), assim como, a maioria dos alunos não reside na casa de estudante universitário (96%). A análise de desempenho acadêmico mostra que a média geral de notas dos alunos está em 5,472 apresentando um desvio padrão de 2,414. A análise da posição do aluno considera a carga horária do curso, ou seja, a análise da carga horária cumprida pelo aluno e o semestre cronológico que o aluno se encontra atualmente no curso. Portanto, esse atributo analisa se o aluno está em dia com o que dele se espera, evitando-se assim a retenção acadêmica. O valor médio desse atributo é -2,548 e varia de -14 a +4 entre os 394 alunos do curso.

Para a fase de modelagem foram utilizados vários métodos existentes na literatura (QUINLAN, 1993; AHA *et al.*, 1991; BREIMAN *et al.*, 1984; JOHN & LANGLEY, 1995; HOLTE, 1993; HORNIK *et al.*, 1989). Cada simulação é composta por um conjunto de modelos. Além disso, em uma mesma simulação, um determinado modelo pode ser aplicado várias vezes com configurações e parametrizações diferentes. Os modelos aplicados na etapa de modelagem foram C4.5 (QUINLAN, 1993), *nearest neighbor* (AHA *et al.*, 1991), CART (BREIMAN *et al.*, 1984), *Multilayer Perceptron* (HORNIK *et al.*, 1989), *Naïve Bayes* (JOHN & LANGLEY, 1995) e *OneR* (HOLTE, 1993). Para fins de análise, os modelos foram aplicados em todas as matrículas realizadas no primeiro semestre de 2015, totalizando 302 alunos regulares no curso de Zootecnia.

Todas as simulações trabalharam com atributos normalizados e não normalizados. A simulação 1 utilizou cinco modelos (AHA *et al.*, 1991; BREIMAN *et al.*, 1984; QUINLAN, 1993; JOHN & LANGLEY, 1995; HORNIK *et al.*, 1989), mas somente um (AHA *et al.*, 1991) apresentou resultados mensuráveis. A Tabela 1 apresenta os resultados encontrados para a simulação 1. A primeira coluna apresenta o modelo utilizado, a segunda coluna descreve se os atributos foram normalizados e a terceira coluna mostra o número total previsto e o percentual da previsão no total analisado. Os valores de  $k$  na primeira coluna da Tabela 1 representam o número de instâncias que serão comparadas para a previsão. Por exemplo, se  $k=3$  serão analisadas as três instâncias mais próximas para determinar a previsão. Os valores previstos de potenciais evasões variam de 1 a 30 e não são exclusivos. Existe uma sobreposição no total previsto. A previsão obteve 14 matrículas presentes em quatro técnicas, quatro matrículas em três técnicas, oito matrículas em duas técnicas e, finalmente, 12 matrículas em um modelo. A relação final, considerando todos os modelos, possuem 38 matrículas, o que representa 12,5% do total de matriculados.

**Tabela 1. Previsão da Simulação 1**

Modelo	Atributos normalizados	Previsão
AHA, 1991 – $k=1$	Não	2 (0,6%)
AHA, 1991 – $k=3$	Não	26 (9%)
AHA, 1991 – $k=5$	Não	30 (10%)
AHA, 1991 – $k=1$	Sim	1 (0,3%)
AHA, 1991 – $k=3$	Sim	19 (6%)
AHA, 1991 – $k=5$	Sim	18 (6%)

A simulação 2 foi aplicada nos mesmos cinco modelos da simulação 1 porém os resultados alcançados foram diferentes. Entre os cinco modelos aplicados, três (AHA *et al.*, 1991; HORNIK *et al.*, 1989; QUINLAN, 1993) produziram resultados interessantes. A Tabela 2 ilustra os resultados alcançados. O modelo *nearest neighbor* (AHA *et al.*, 1991) foi aplicado com três configurações diferentes. Os valores previstos nessa simulação variam de 1 a 11 e, também, não são exclusivos. Os experimentos mostraram uma relação de 15 matrículas diferentes entre os três modelos aplicados. O detalhamento da previsão mostra que uma matrícula foi prevista em cinco modelos, seis matrículas foram previstas em quatro modelos, cinco matrículas em dois modelos e três matrículas em somente em um modelo. A simulação 2 obteve um número menor de previsões porque atributos que consideram o desempenho acadêmico foram adicionados ao conjunto de treinamento e teste. Mais especificamente, atributos tais como número de disciplinas reprovadas, trancamentos de disciplinas, trancamentos de semestres e disciplinas reprovadas por frequência.

Finalmente, na simulação 3 foram adicionados dois novos atributos. Os atributos posição no período e desempenho acadêmico (sequências 19 e 28 no Apêndice I, respectivamente). A Tabela 2 descreve os resultados obtidos com a simulação 3. Na simulação 3 foram aplicados vários modelos (QUINLAN, 1993; AHA *et al.*, 1991; BREIMAN *et al.*, 1984; JOHN & LANGLEY, 1995; HOLTE, 1993; HORNIK *et al.*, 1989), mas somente dois (QUINLAN, 1993; BREIMAN *et al.*, 1984) apresentaram resultados interessantes. O modelo C4.5 (QUINLAN, 1993) foi utilizado com duas configurações diferentes. A primeira configuração utiliza árvores de decisão não binárias, enquanto que a segunda configuração gera o modelo com árvores binárias. De maneira

similar às demais simulações, atributos normalizados e não normalizados foram usados. O número total previsto para a simulação 3 foi de 37, e a variação dos modelos ficou entre 4 e 23. Entre as 37 previsões, uma delas foi prevista em seis modelos, uma em cinco modelos, duas em quatro modelos, oito em três modelos, seis em dois modelos e 19 em somente um modelo.

**Tabela 2. Previsão da Simulação 2 e Simulação 3**

Previsão da Simulação 2			Previsão da Simulação 3		
Modelo	Atributos Normalizados	Previsão	Modelo	Atributos Normalizados	Previsão
QUINLAN, 1993	Não	3 (1%)	QUINLAN, 1993 – Binária	Não	5 (2%)
AHA, 1991 – k=3	Não	10 (3%)	QUINLAN, 1993	Não	4 (1%)
AHA, 1991 – k=5	Não	8 (3%)	BREIMAN, 1984	Não	4 (1%)
AHA, 1991 – k=3	Sim	11 (4%)	QUINLAN, 1993 – Binária	Sim	23 (8%)
AHA, 1991 – k=5	Sim	9 (3%)	QUINLAN, 1993	Sim	22 (7%)
HORNIK, 1989	Sim	1 (0,3%)	BREIMAN, 1984	Sim	16 (5%)

A avaliação das três simulações permite uma visão holística e uma visão cartesiana das previsões. A visão holística ilustra todas as simulações em um modelo único e integrado. A visão cartesiana estuda cada modelo como uma coleção de partes dissociadas. O viés cartesiano visualiza três simulações isoladas, em que a primeira simulação é composta por 38 matrículas, a simulação II é formada por 15 matrículas e, finalmente, a simulação III é composta de 37 matrículas. Por sua vez, o viés holístico mostra uma visão interconectada das simulações, na qual a totalidade das matrículas não se dá pela simples soma das simulações. Esta visão integrada apresenta uma relação de 65 matrículas possíveis de evasão.

## 5. Análise da Previsão – Resultados

Esta seção apresenta os resultados encontrados após o processo de matrícula ser realizado na Instituição. O número de alunos regulares matriculados no período 2015/1 foi 302. Dos 302 alunos regulares matriculados, os alunos que ingressaram no semestre corrente, no total de 36 alunos, foram desconsiderados para previsão. Esse descarte foi realizado, pois os alunos ingressantes ainda possuem um perfil comportamental que permita a previsão de possível evasão.

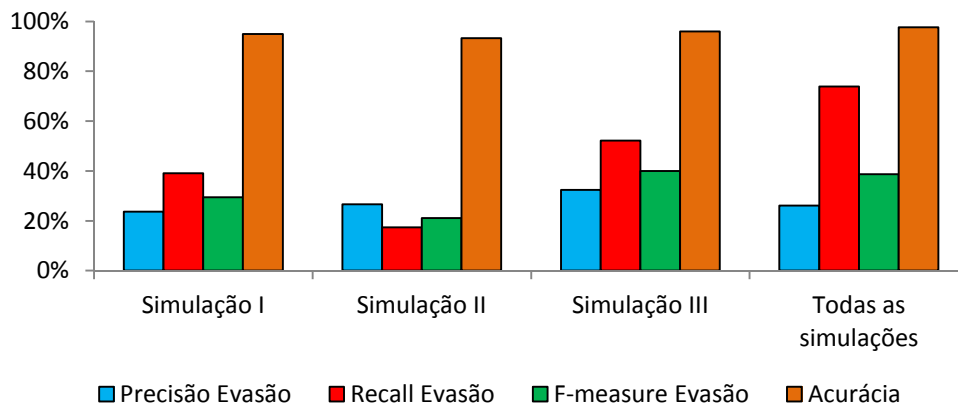
O número de alunos aptos a realizar a matrícula no semestre de 2015/2 foram 299 alunos. Após o prazo de realização de matrículas foram identificados 23 abandonos no curso de Zootecnia. A mineração dos dados acertou 17 dos 23 alunos que abandonaram o curso. Os 17 alunos estavam na relação de 65 matrículas, as quais foram previstas pela consolidação das três simulações.

Para avaliação dos resultados da previsão foram utilizadas as métricas de acurácia, precisão, *recall* e *f-measure*. Essas métricas foram utilizadas para avaliar a qualidade da previsão. A acurácia mede o total de acerto comparado com o total de instâncias avaliadas. A precisão avalia a fração relevante comparada com o total previsto. O *recall* avalia a fração que foi prevista comparada com o total relevante. A métrica *f-measure* combina a precisão e o *recall* em uma só medida. Seja  $R$  uma relação de todas as matrículas previstas. Seja  $S$  um conjunto de todas as matrículas que evadiram durante o processo de matrícula. As definições de precisão ( $P$ ), *recall* ( $R$ ) e *f-measure* ( $F$ ) são dadas por:  $P = \frac{|R \cap S|}{|R|}$ ,  $R = \frac{|R \cap S|}{|S|}$ ,  $F = 2 \times \frac{P \times R}{P + R}$ , respectivamente.

A avaliação foi realizada para cada simulação e também com a combinação das três simulações. A Figura 4 apresenta os valores alcançados para cada uma das métricas em cada simulação e na combinação das simulações. A Figura 4 mostra algumas informações interessantes. Todas as simulações alcançaram um valor alto de acurácia. Isso aconteceu porque a maioria dos alunos foi prevista como alunos regulares. O valor máximo de acurácia foi de 98% e a média entre as simulações alcançou 95%. As simulações I e II apresentaram resultados inferiores comparados



com a simulação III e a combinação das três simulações. A precisão teve o melhor desempenho na simulação III, alcançando o valor percentual de 33%. O *recall*, por sua vez, teve o melhor desempenho na combinação das três simulações, com o valor percentual de 74%. A *f-measure* alcançou o melhor resultado para a simulação III totalizando 0,4. A combinação das simulações teve resultado semelhante (0,39). A *f-measure* obteve o melhor resultado para a simulação III considerando-se que a precisão e o *recall* possuem a mesma importância. Caso o objetivo seja acertar mais alunos que evadem, a combinação das três simulações é melhor que a simulação III.



**Figura 4. Métricas de Avaliação**

## 6. Conclusão

Este artigo propõe uma nova abordagem para previsão de possíveis evasões de cursos de graduação presenciais em Universidades públicas. A abordagem extrai informações pessoais, acadêmicas, sociais e econômicas e, por meio das informações extraídas, são construídos modelos de previsão utilizando-se métodos de mineração de dados. A vantagem da abordagem proposta é a combinação de vários modelos de mineração de dados em uma única previsão, otimizando o resultado do processo. Essa visão holística, na qual os modelos são combinados em um todo único e integrado, permite um resultado mais abrangente e preciso. Os experimentos demonstraram que a abordagem proposta nesse trabalho é apropriada para prever a evasão no curso de graduação em Zootecnia. As métricas alcançaram 98% na previsão das classes envolvidas (aluno regular e aluno evadido) e uma taxa de sucesso em 74% na previsão dos alunos que abandonaram o curso.

Como trabalhos futuros pretende-se realizar o estudo dos atributos que determinaram o sucesso do processo para filtrar somente aqueles que interessam à previsão. Dessa forma, são identificados os principais fatores que levaram ao processo de evasão. Além disso, a coleta de novos atributos pode ser realizada pela mineração em redes sociais (por exemplo, *facebook*, *twitter*, entre outras) e redes profissionais (por exemplo, *linkedin*), as quais contêm informações sobre as atividades comportamentais dos alunos. Outra perspectiva tem como foco a generalização desse processo para os demais cursos de graduação, uma vez que o perfil de comportamento dos alunos em diferentes cursos é distinto. Essa generalização pode ser alcançada por meio de automatização do processo de previsão, nos quais o especialista analisa o resultado das previsões para apoiar ações de mitigação no processo de evasão. Espera-se que esse processo contribua na valorização de novas ações no sentido de minimizar a evasão de alunos dos cursos de graduação. Acredita-se que a possibilidade de gerar uma lista menor de potenciais alunos, permita ao curso realizar uma melhor gestão de permanência do aluno junto ao curso.

## Referências

ABU-ODA, Ghadeer S.; EL-HALEES, Alaa M. DATA MINING IN HIGHER EDUCATION: UNIVERSITY STUDENT DROPOUT CASE STUDY. *International Journal of Data Mining & Knowledge Management Process*, v. 5, n. 1, p. 15, 2015.



- ADACHI, A. A. C. T. Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais. **Belo Horizonte: UFMG/FaE, (Dissertação de Mestrado)**, 2009.
- AGICHTEIN, Eugene; BRILL, E.; DUMAIS, S.; RAGNO, R.. Learning user interaction models for predicting web search result preferences. In: **Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval**. ACM, 2006. p. 3-10.
- AHA, David W.; KIBLER, Dennis; ALBERT, Marc K. Instance-based learning algorithms. **Machine learning**, v. 6, n. 1, p. 37-66, 1991.
- AMARAL, João Batista do. Evasão discente no ensino superior: estudo de caso no Instituto Federal de Educação, Ciência e Tecnologia do Ceará (Campus Sobral). 2013.
- AMAYA TORRADO, Yegny Karina; BARRIENTOS AVENDAÑO, E.; HEREDIA VIZCAÍNO, D.J. Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. 2014.
- AZOUMANA, Kamagate. Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas. **Revista Pensamiento Americano**, v. 6, n. 10, 2014.
- BRAUSE, R.; LANGSDORF, T.; HEPP, Michael. Neural data mining for credit card fraud detection. In: **Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on**. IEEE, 1999. p. 103-106.
- BREIMAN, Leo; FRIEDMAN, J.; STONE, C.J.; OLSHEN, R.A. **Classification and regression trees**. CRC press, 1984.
- CHAN, Philip K.; FAN, W.; PRODRONIDIS, A.L.; STOLFO, S.J. Distributed data mining in credit card fraud detection. **Intelligent Systems and their Applications, IEEE**, v. 14, n. 6, p. 67-74, 1999.
- CHAN, Philip K.; STOLFO, Salvatore J. Toward Scalable Learning with Non-Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. In: **KDD**. 1998. p. 164-168.
- CHAPMAN, Pete; CLINTON, J.; KERBER, R.; KHAHAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. CRISP-DM 1.0 Step-by-step data mining guide. 2000.
- FAYYAD, Usama; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.
- GAIOSO, N. P. de L. **O Fenômeno da evasão escolar na educação superior no Brasil**. Brasília: 2005.
- HOLTE, Robert C. Very simple classification rules perform well on most commonly used datasets. **Machine learning**, v. 11, n. 1, p. 63-90, 1993.
- HORNIK, Kurt; STINCHCOMBE, Maxwell; WHITE, Halbert. Multilayer feedforward networks are universal approximators. **Neural networks**, v. 2, n. 5, p. 359-366, 1989.
- JADRIĆ, Mario; GARAČA, Željko; ČUKUŠIĆ, Maja. Student dropout analysis with application of data mining methods. **Management: J. of Contemporary Management Issues**, v. 15, n. 1, p. 31-46, 2010.
- JOHN, George H.; LANGLEY, Pat. Estimating continuous distributions in Bayesian classifiers. In: **Proceedings of the Eleventh conference on Uncertainty in artificial intelligence**. Morgan Kaufmann Publishers Inc., 1995. p. 338-345.
- KANTARDZIC, Mehmed. Data-Mining Concepts. **Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition**, p. 1-25, 2011.
- KANTORSKI, G.Z.; MORAES, T.G.; MOREIRA, V.P.; HEUSER, C.A. Choosing Values for Text Fields in Web Forms. In: **Advances in Databases and Information Systems**. Springer Berlin, 2013. p. 125-136.
- KANTORSKI, G. Z.; HEUSER, C. A. Automatic Filling of Web Forms. In: **AMW**. 2012. p. 215-219.
- LEAL, R.; MOTEJUNAS, P. R., HIPÓLITO, O., LOBO, M. B. V. M. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, vol.9, n.132, pp. 641-659, set/dez. 2007.

- LOBO, M. B. de C. M. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. **Associação Brasileira de Mantenedoras de Ensino Superior. Cadernos**, n. 25, 2012.
- MEC. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Revista de rede de avaliação institucional da educação superior. Campinas, v. 1, n. 2, p. 55-65, dez.1996.
- MEC/SESU. Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Brasília: ANDIFES/ABRUEM/SESU/MEC. 1997.
- MITCHELL, Tom M. Machine learning. 1997. **Burr Ridge, IL: McGraw Hill**, v. 45, 1997.
- MOROSINI, M. C.; CASARTELLI, A.D.O.; SCHIMITT, R.E.; SANTOS, B.S.; BENSO, A.C.; GESSINGER, R.M. **A Evasão na Educação Superior no Brasil: uma análise da produção de conhecimento nos periódicos**. PUCRS. 10 f. p.1-10, 2011.
- OREA, Sergio Valero; VARGAS, Alejandro Salvador; ALONSO, Marcela García. Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. **Ene**, v. 779, n. 73, p. 33, 2005.
- PALÁCIO, Paula da Paz. Políticas de acesso e permanência do estudante da Universidade Federal do Ceará (UFC). 123p. Dissertação. UFC, Fortaleza, 2012.
- PEREIRA JR, E. **Compromisso com o graduar-se, com a instituição e com o curso: estrutura fatorial e relação com a evasão**. 2012. 414p. Dissertação. Universidade Estadual de Campinas, Campinas, SP, 2012.
- POMERLEAU, D.; JOCHEM, T. (1996). Rapidly adapting machine vision for automated vehicle steering. *IEEE Intelligent Systems*, (2), 19-27.
- QUINLAN, John Ross. **C4. 5: programs for machine learning**. Morgan kaufmann, 1993.
- SCALI, Danyelle Freitas. **Evasão nos Cursos Superiores de Tecnologia: a Percepção dos Estudantes sobre seus Determinantes**. 2009.140 f. Dissertação (Mestrado em Educação) Faculdade de Educação. São Paulo: Universidade Estadual de Campinas, 2009.
- SILVA FILHO, Roberto Leal Lobo; MONTEJUNAS, P.R.; HIPÓLITO, O. LOBO, M.B.C.M. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, n. 132, p. 641-659, 2007.
- SOBRINHO, José Dias. Avaliação educativa: produção de sentidos com valor de formação. **Avaliação, Campinas**, v. 13, n. 1, p. 193-207, 2008.
- TIGRINHO, Luiz Maurício Valente. Evasão escolar nas instituições de ensino superior. **Revista Gestão Universitária**, v. 173, p. 01-14, 2008.
- TINTO, V. Dropout from higher education: a theoretical synthesis of recent research. **Review of Educational Research**, 45: 89-125. 1975.
- TINTO, Vincent. Limits of theory and practice in student attrition. **The journal of higher education**, p. 687-700, 1982.
- TONTINI, Gérson et al. PODE-SE IDENTIFICAR A PROPENSÃO E REDUZIR A EVASÃO DE ALUNOS? AÇÕES ESTRATÉGICAS E RESULTADOS TÁTICOS PARA INSTITUIÇÕES DE ENSINO SUPERIOR. **Avaliação: Revista da Avaliação da Educação Superior**, v. 19, n. 1, 2013.