

## Predição da Evasão em Cursos de Graduação em Instituições Públicas

Gustavo Zanini Kantorski<sup>1</sup>, Evandro Gomes Flores<sup>1</sup>, Ivan Londero Hoffmann<sup>1</sup>,  
Jáder Adiel Schmitt<sup>1</sup>, Fernando Pires Barbosa<sup>1</sup>

<sup>1</sup>Universidade Federal de Santa Maria (UFSM)  
97.105-900 – Santa Maria – RS – Brasil

**Abstract.** *This paper presents the college dropout that exists on undergraduate programs of public Institutes in Brazil. The aim is the dropout prediction in undergraduate programs, which helps visualize perspectives of motivation, reducing the dropout process. In this paper, we propose several machine learning methods for prediction. The intuition is to generate a list of likely students, which will be not enrolling in courses in next period. The experiments demonstrate accuracy around 98% in general prediction, and dropout prediction rate around 70%.*

**Resumo.** *Este artigo aborda a evasão existente na 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 de aprendizagem de máquina para a previsão. A intuição é gerar uma lista de prováveis alunos que não realizarão matrícula no período posterior ao que estão cursando. Os resultados demonstram que a abordagem proposta é factível e eficiente. Os experimentos alcançaram uma acurácia de 98% na previsão, e mais de 70% de sucesso na previsão de alunos que abandonaram o curso.*

### 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 e econômicas; e, também, para a família, pelo sonho não realizado [Silva Filho et al. 2007].

A maioria dos trabalhos que trata a evasão no ensino superior tem como finalidade compreender os fatores que causam a evasão [Amaral 2013, Rodrigues et al. 2015, Reino et al. 2015, Palácio 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.

Existem trabalhos que abordam a previsão de determinadas situações do processo educacional. Por exemplo, trabalhos que tratam a previsão de desempenho de estudantes [Costa et al. 2015, Adeodato et al. 2014, de Brito et al. 2014], predição de reprovação em disciplinas [Detoni et al. 2014] e que abordam a evasão no sentido de prever situações de evasão no ensino superior [Silva et al. 2015, Abu-Oda and El-Halees 2015, Jadrić et al. 2010, Amaya Torrado et al. 2014]. Geralmente, esses trabalhos utilizam técnicas de descoberta de conhecimento em bases de dados ou mineração de dados para realizar a previsão de resultados.

O presente trabalho apresenta uma abordagem para a previsão da evasão em cursos de graduação presenciais. A intuição é realizar uma análise comportamental dos alunos 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.

Este artigo está organizado da seguinte forma. A Seção 2 apresenta a metodologia desenvolvida para análise da evasão. A Seção 3 descreve a avaliação experimental realizada, detalhando as características extraídas para geração dos modelos preditivos, os modelos utilizados para análise e a previsão dos modelos em uma visão holística e cartesiana. A Seção 4 apresenta os resultados e os experimentos realizados. Na Seção 5 são apresentados os trabalhos relacionados. Considerações finais e direções futuras são discutidos na seção 6.

## 2. Metodologia

A metodologia proposta utiliza conceitos de aprendizagem de máquina para prever as informações relacionadas à evasão. Para análise da evasão 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 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.

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. Dentre as atividades de

limpeza pode-se citar a eliminação de dados com erro, a padronização de dados, por exemplo, abreviaturas, formatos de data, valores de atributos. 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 integração serve para combinar múltiplas fontes de dados e a formatação pode realizar alterações sintáticas nos dados sem modificar o seu significado.

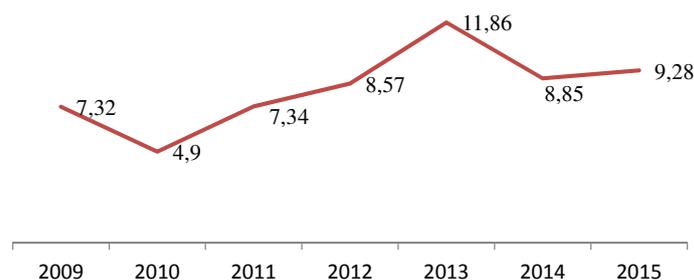
A etapa de modelagem é a fase de descoberta de conhecimento propriamente dita. Nessa etapa são selecionados os modelos que serão aplicados para a solução da tarefa. A etapa de avaliação interpreta e avalia os resultados em relação aos objetivos de descoberta de conhecimento definidos na etapa de entendimento 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.

### 3. Avaliação Experimental

A metodologia apresentada na seção 2 foi aplicada a dois cursos de graduação presencial da Instituição: Curso de Zootecnia e Curso de Administração. Por questões de espaço, serão abordados os resultados de um dos cursos - Administração. Para o cálculo da evasão foi utilizada a fórmula definida por [Silva Filho et al. 2007], a qual é baseada 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, dada por  $E_n = 1 - \frac{M_n - I_n}{M_{n-1} - C_{n-1}}$ , em que  $E_n$  é a taxa de evasão anual,  $M$  é o número de matriculados,  $I$  representa o número de ingressantes,  $C$  o número de concluintes e  $n$  o ano em estudo. 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 1 apresenta a taxa de evasão do curso de Administração, desde 2009 até 2015 e o valor médio da taxa de evasão é 8,3 nos últimos sete anos.

A metodologia proposta foi aplicada em duas simulações. Uma simulação possui um conjunto de atributos e um conjunto de modelos. O que diferencia uma simulação de outra são o número de instâncias de treinamento e os algoritmos de descoberta de conhecimento aplicados a esses atributos. As duas simulações possuem 32 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*. A diferença entre as simulações é que a primeira simulação considera o aluno Formado como se fosse um Aluno Regular enquanto que a segunda simulação não faz essa consideração. Além disso, em uma mesma simulação, foram utilizadas técnicas para normalização dos atributos numéricos. 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.

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 [Silva Filho et al. 2007, Sobrinho 2008], existentes na literatura sobre o tema. Os atributos coletados foram os seguintes: sexo, idade, etnia, estado civil, cotista (sim/não), formação no ensino médio, já iniciou curso superior



**Figura 1. Taxa de Evasão para o Curso de Administração**

(sim/não), como vai se manter durante o curso, motivo que optou pelo curso, bolsa especial da instituição (BSE), atendimento psicossocial (sim/não), ano de ingresso, tempo no curso, período atual (período ideal que o aluno está no curso), número de disciplinas aprovadas, reprovadas, reprovadas por frequência, dispensadas, trancamentos totais, trancamentos parciais, média geral das notas, total de disciplinas matriculadas, média de disciplinas por período, moradia estudantil (S/N), monitoria (S/N), média de utilização do Restaurante Universitário, média de utilização da biblioteca, posição no período (razão entre carga horária total do curso e o número de períodos que o aluno deve cursar), desempenho acadêmico. 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 curso de Administração de 2000 a 2015 foram consideradas para análise. No período indicado foram selecionadas 791 instâncias de treinamento para a simulação I. Do ponto de vista da categoria pessoal, do total de instâncias de treinamento, 407 são do sexo feminino e 384 do sexo masculino. Ainda, 673 instâncias não ingressaram por sistema de cotas, enquanto que 118 cotistas estão no conjunto. A análise sócio econômica mostra que a maioria dos alunos é proveniente de escola pública (364), assim como a maioria dos alunos irá se manter com recursos familiares durante o curso (724). A maioria das informações da categoria Institucional é faltante (em torno de 93%). Isso ocorre porque essas informações são geradas no processo de avaliação institucional, em que a maioria dos alunos não participa ativamente. Considerando a simulação II tem-se 339 instâncias de treinamento. Na categoria pessoal, do total de instâncias de treinamento, 177 instâncias do sexo masculino e 162 do sexo feminino. Ainda, 251 instâncias não ingressaram por sistema de cotas, enquanto que 88 são cotistas. Os demais atributos possuem comportamento similar àqueles apresentados na simulação I.

O viés de assistência estudantil descreve que existem mais alunos bolsistas (61%) na simulação I que na simulação II (34%). Na perspectiva de bolsas de monitoria a situação é inversa em ambas as simulações. A maioria dos alunos não possui bolsa de monitoria (acima de 95%), assim como, a maioria dos alunos não reside na casa de estudante universitário (99%). A análise de desempenho acadêmico mostra que a média geral de notas dos alunos está em 5,6 na simulação II e 6,7 na simulação I. 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. Por-

**Tabela 1. Previsão das Simulações**

Algoritmo	Simulação I		Simulação II	
	Atributos Não Normalizados	Atributos Normalizados	Atributos Não Normalizados	Atributos Normalizados
<i>CART</i>	03/02	23/06	07/04	35/10
<i>J48</i>	05/02	29/09	07/04	32/10
<i>J48-Bin</i>	02/01	23/06	04/02	28/08
<i>Naïve Bayes</i>	23/08	N/A	10/04	N/A

tanto, 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 -6,56 e varia de -26 a +3 entre os 791 alunos presentes na simulação I e de -4 a -26 entre os 339 alunos da simulação II.

Para a fase de modelagem foi utilizada a ferramenta Weka [Hall et al. 2009]. Essa ferramenta consiste de um conjunto de algoritmos e modelos de aprendizagem de máquina. 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 algoritmos aplicados na etapa de modelagem foram J48, nearest neighbor (IBk), CART, Multilayer Perceptron, Naïve Bayes e OneR. Para fins de análise, os modelos foram aplicados nos dados de todos os alunos aptos a realizar matrícula no primeiro semestre de 2016, totalizando 201 alunos regulares no curso de Administração.

Todas as simulações trabalharam com atributos normalizados e não normalizados. A simulação I utilizou cinco algoritmos (J48, IBk, CART, Naïve Bayes, Multilayer Perceptron), mas somente três (J48, CART e Naïve Bayes) apresentaram resultados mensuráveis. A Tabela 1 apresenta os resultados encontrados para a simulação I. O algoritmo J48 foi aplicado com duas configurações. A opção *Bin*, na Tabela 1, identifica que o algoritmo utilizou árvores binárias para criar o modelo. A primeira coluna apresenta o modelo utilizado, a segunda e quarta colunas descrevem que os atributos não estão normalizados e a terceira e quinta colunas mostram os atributos normalizados. O valor previsto no conjunto de teste é apresentado no numerador da relação, enquanto o valor real é apresentado no denominador. Os valores previstos de potenciais evasões variam de 02 a 29 e não são exclusivos. Existe uma sobreposição no total previsto. A previsão obteve uma matrícula presente em todos os modelos da Tabela 1, duas matrículas em seis modelos, três matrículas em cinco modelos, 13 matrículas em quatro modelos, quatro matrículas em três modelos e, finalmente, 10 matrículas em um modelo. A relação final, considerando a união de todos os modelos, possuem 33 matrículas.

A simulação II foi aplicada nos mesmos cinco modelos da simulação I porém os resultados alcançados foram diferentes. A Tabela 1 ilustra os resultados alcançados. O modelo J48 foi aplicado com duas configurações diferentes. Os valores previstos nessa simulação variam de 04 a 35 e, também, não são exclusivos. O detalhamento da previsão mostra que quatro matrículas foram previstas em sete modelos, duas matrículas foram previstas em cinco modelos, três matrículas em quatro modelos, 22 matrículas em três modelos, uma matrícula em dois modelos e cinco matrículas em somente um modelo. A simulação II obteve um número maior de previsões, alcançando o resultado de 39 matrículas.

A avaliação das duas 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 duas simulações isoladas, em que a primeira simulação é composta por 33 matrículas e a simulação II é formada por 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 37 matrículas possíveis de evasão.

#### 4. Resultados

Esta seção apresenta os resultados encontrados após o processo de matrícula ser realizado na Instituição. O número de alunos aptos a realizar a matrícula no semestre de 2016/1 foram 201 alunos. Após o prazo de realização de matrículas foram identificados 15 abandonos no curso de Administração. A mineração dos dados acertou 11 dos 15 alunos que abandonaram o curso. Os 11 alunos estavam na relação de 39 matrículas, as quais foram previstas pela consolidação das duas simulações realizadas.

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 duas simulações. A Figura 2 apresenta os valores alcançados para cada uma das métricas em cada simulação e na combinação das simulações. A Figura 2 mostra algumas informações interessantes. A simulação I foi superior em relação à simulação II em todas as métricas e superior a união das simulações em duas métricas (precisão e *f-measure*). A precisão teve o melhor desempenho na simulação I, alcançando o valor de 33%. O *recall* teve o melhor desempenho na simulação I e na união das simulações, com o valor de 73%. A *f-measure* alcançou o melhor resultado para a simulação I totalizando 46%. A combinação das simulações teve resultado semelhante (41%).

A Figura 3 ilustra um gráfico detalhado para cada método de mineração aplicado em cada simulação. As métricas de precisão, *recall* e *f-measure* são apresentadas para a classe de evasão. A métrica de *f-measure* é apresentada em três formatos. O primeiro, *f-measure*, considera a precisão e o *recall* com o mesmo peso. O segundo, *f-measure - recall*, considera o *recall* com o dobro do peso que a precisão e o terceiro, *f-measure - precisao*, enfatiza a precisão com o dobro do peso do *recall*. Ainda, as letras S e N significam que os dados estão ou não estão normalizados, respectivamente.

O gráfico da Figura 3 demonstra que o modelo CART aplicado em dados não normalizados, na simulação I, apresentou o melhor resultado para a métrica de precisão enquanto que os modelos CART e J48, com dados normalizados, na simulação II, apresentaram o melhor resultado para o *recall*. Considerando a *f-measure*, a melhor técnica de mineração foi a CART aplicada a dados normalizados. O gráfico destaca a importância

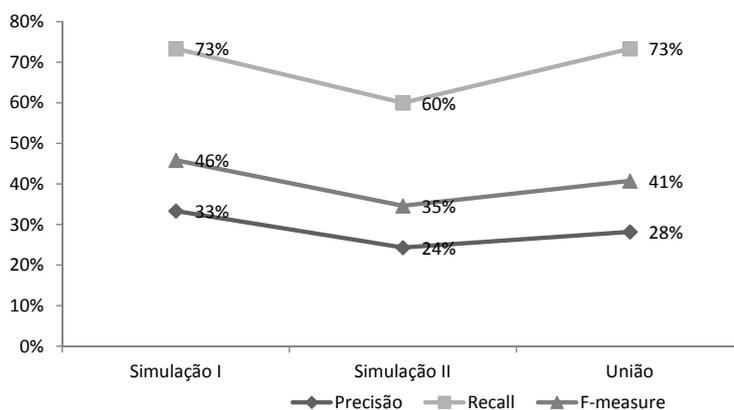


Figura 2. Métricas de Avaliação

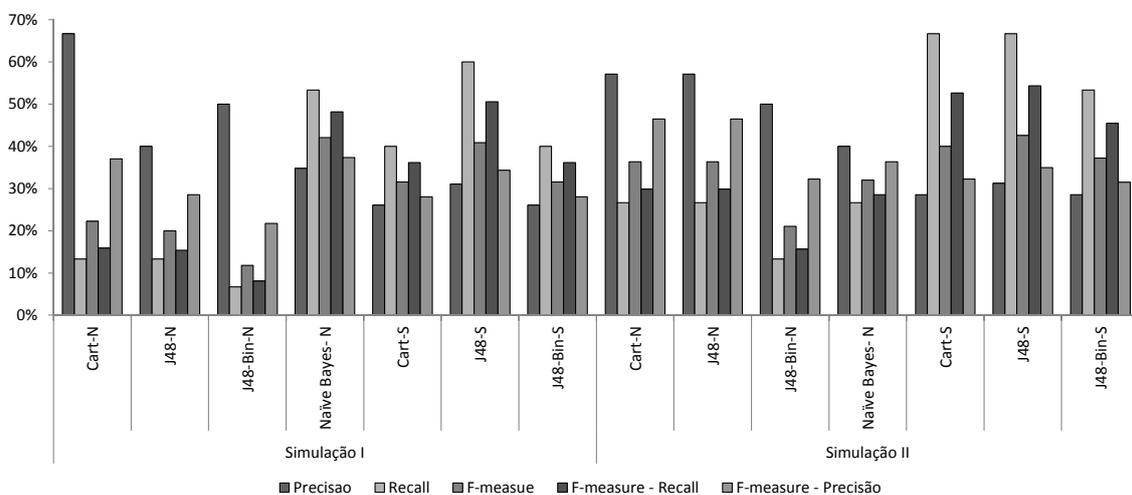


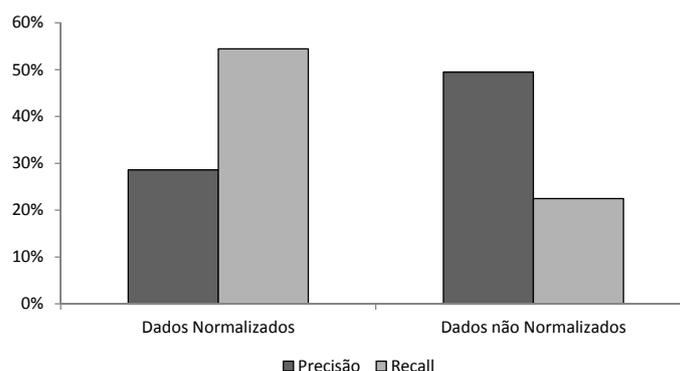
Figura 3. Métricas de Avaliação

da manipulação dos dados na mineração. Pode-se observar para a técnica CART com dados normalizados obteve um desempenho duas vezes melhor que a CART com dados não normalizados. Isso demonstra a importância da etapa de preparação dos dados na metodologia apresentada na seção 2. Considerando a métrica de *recall*, a técnica CART também foi superior entre todas as demais técnicas, alcançando o valor de 35%. A melhor precisão também foi superior na simulação I, com a técnica CART, atingindo 67%, mas a medida de *recall* foi muito baixa (13%), o que implica em um resultado não satisfatório.

A Figura 4 apresenta uma análise das métricas de precisão e *recall* sob o ponto de vista da normalização dos dados. Independente das simulações, oito configurações de técnicas de mineração foram aplicadas com os dados não normalizados e seis configurações de técnicas de mineração foram executadas com dados normalizados. A métrica de precisão alcançou uma média de 49% de acerto para dados não normalizados em relação ao valor de 28% para dados normalizados. Para a métrica de *recall* o valor para dados normalizados foi de 54% em relação aos 22% para dados não normalizados.

Observando as Figuras 2 e 3 é possível verificar que o cálculo da média dos valores das métricas de precisão, *recall* e *f-measure* não combinam entre as figuras, isto é, a média das métricas de cada simulação, na Figura 4, não, necessariamente, é igual a média das

técnicas de mineração associadas à simulação correspondente na Figura 2. Isso acontece porque a previsão não pode ser baseada na soma das métricas e no cálculo da média geral. Ao observar a Figura 4, é possível verificar que a combinação das simulações não é, meramente, a soma das simulações I e II. Portanto, a visão holística do método não é linear quando as técnicas de mineração são combinadas. Além disso, a combinação das simulações alcança valores superiores comparados com os valores de cada simulação e com a média das simulações.



**Figura 4. Dados Normalizados x Dados Não Normalizados**

## 5. Trabalhos Relacionados

Existem vários trabalhos relacionados que tratam a questão da evasão. O trabalho proposto por [Amorim et al. 2008] concentra-se na modelagem dos principais aspectos que levam o aluno que ainda não concluiu seu curso, a abandoná-lo ou trancá-lo. O trabalho utiliza três classificadores e leva em consideração quatro cursos para análise. O valor médio de acurácia varia entre 89% e 91%. [Manhães et al. 2011] propõem a identificação da evasão por meio das notas e situação final dos alunos nas disciplinas mais cursadas relativas ao primeiro período do curso. O trabalho realizou três experimentos e o valor médio de acurácia variou entre 75% a 80%. O trabalho utiliza um curso e 10 algoritmos de classificação. A abordagem proposta por [Detoni et al. 2014] aplica técnicas de aprendizado de máquina para prever o risco de reprovação de um estudante de educação a distância utilizando como atributos unicamente contagens de interações. Os autores aplicam as técnicas em dois cursos e utilizam quatro modelos de aprendizagem. A acurácia média dos classificadores foi de 90%. A precisão dos alunos reprovados manteve-se entre 75% e 95%. Predições de desempenho e sucesso também fazem parte de trabalhos relacionados. O trabalho proposto por [Costa et al. 2015] ilustra um método para identificação de estudantes cotistas com possibilidade de ter menor rendimento acadêmico no decorrer do curso. Dados de alunos de graduação cotistas da UFPB são utilizados para realizar a avaliação. O método utiliza quatro atributos de entrada e foram utilizados cinco classificadores para avaliação. A acurácia dos modelos varia de 67% a 88%. Outro trabalho de predição de desempenho proposto por [de Brito et al. 2014] aponta para o desempenho dos alunos do primeiro período do curso de Ciência da Computação da UFPB, por intermédio das suas notas de ingresso no processo seletivo. O trabalho utiliza notas de ingresso e de quatro disciplinas do primeiro período do curso. A análise utiliza cinco classificadores para avaliação dos resultados. A acurácia alcançada é superior a 70%. O nosso trabalho difere dos demais pois não somente utiliza os algoritmos de aprendizagem de máquina para a previsão, mas realiza a combinação do resultado de vários modelos de aprendizagem para avaliação da previsão. Esse viés holístico de combinar os modelos permite uma visão mais integrada dos diversos modelos gerados na previsão.

## 6. Conclusão

Este artigo propõe uma abordagem para previsão de possíveis evasões de cursos de graduação presenciais em Universidades públicas. A abordagem utiliza conceitos de aprendizagem de máquina e a tarefa supervisionada de classificação. O método proposto extrai informações pessoais, acadêmicas, sociais e econômicas e, por intermédio das informações extraídas, são construídos modelos de previsão utilizando-se algoritmos de aprendizagem de máquina. A vantagem da 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 Administração da Instituição. As métricas alcançaram 95% na previsão das classes envolvidas (aluno regular e aluno evadido) e uma taxa de sucesso em 73% na previsão dos alunos que não solicitaram matrícula no curso. A metodologia foi aplicada também para o curso de Zootecnia e apresentou resultados semelhantes, alcançando uma taxa de 74% de sucesso na previsão dos alunos que abandonaram o curso. Outra contribuição é a possibilidade de avaliação dos modelos, a qual permite identificar os principais atributos que ajudam na predição. Essa identificação de atributos influi diretamente nos fatores associados ao processo de evasão e servem de estudo para mitigar o abandono nos cursos de graduação.

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. Atualmente já estão sendo coletadas informações sobre os ambientes virtuais de aprendizagem (AVA), utilizados como apoio ao ensino presencial. A ideia é aplicar a metodologia proposta em cursos semi-presenciais e à distância, uma vez que a metodologia é genérica. 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 muitas 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. Embora a previsão dos cursos de Administração e Zootecnia alcançaram resultados semelhantes, os atributos utilizados para gerar os modelos de previsão foram diferentes. Um estudo precisa ser realizado e adequado para cada curso de graduação. Para futuras previsões pode-se criar um laço de retroalimentação com as previsões realizadas em períodos anteriores. Dessa forma, novos atributos podem ser adicionados aos atuais para ajudar no processo de previsão. 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, G. S. and El-Halees, A. M. (2015). Data mining in higher education: University student dropout case study. *International Journal of Data Mining & Knowledge Management Process*, 5(1):15.
- Adeodato, P. J., Santos Filho, M. M., and Rodrigues, R. L. (2014). Predição de desempenho de escolas privadas usando o enem como indicador de qualidade escolar. In *Anais do SBIE*, volume 25, page 891.

- Amaral, J. B. d. (2013). Evasão discente no ensino superior: estudo de caso no instituto federal de educação, ciência e tecnologia do ceará (campus sobral).
- Amaya Torrado, Y. K., Barrientos Avendaño, E., Heredia Vizcaíno, D. J., et al. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos.
- Amorim, M. J., Barone, D., and Mansur, A. U. (2008). Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica. *XIX SBIE*, pages 12–14.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). Crisp-dm 1.0 step-by-step data mining guide.
- Costa, F., dos Santos Silva, A. R., de Brito, D. M., and do Rêgo, T. G. (2015). Predição de sucesso de estudantes cotistas utilizando algoritmos de classificação. In *Anais do SBIE*, volume 26, page 997.
- de Brito, D. M., de Almeida Júnior, I. A., Queiroga, E. V., and do Rêgo, T. G. (2014). Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. In *SBIE*, volume 25, page 882.
- Detoni, D., Araujo, R. M., and Cechinel, C. (2014). Predição de reprovação de alunos de educação a distância utilizando contagem de interações. In *SBIE*, volume 25, page 896.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations*, 11(1):10–18.
- Jadrić, M., Garača, Ž., and Čukušić, M. (2010). Student dropout analysis with application of data mining methods. *Management: Journal of Contemporary Management Issues*, 15(1):31–46.
- Manhães, L. M. B., Cruz, S. d., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Anais do XXII SBIE-XVII WIE, Aracaju*.
- Palácio, P. d. P. (2012). *Políticas de acesso e permanência do estudante da Universidade Federal do Ceará (UFC). 2012. 123f.* PhD thesis, Dissertação (Mestrado em Políticas Públicas e Gestão da Educação Superior)- Universidade Federal do Ceará, Fortaleza.
- Reino, L. R. A. C., Hernández-Domínguez, A., Júnior, O. d. G. F., Carvalho, V., Barros, P. A. M., and Braga, M. (2015). Análise das causas da evasão na educação a distância em uma instituição federal de ensino superior. In *Anais do SBIE*, volume 26, page 91.
- Rodrigues, F., Brackmann, C., and Barone, D. A. (2015). Estudo da evasão no curso de ciência da computação da ufrgs. *RBIE*, 23(01):97.
- Silva, F., da Silva, J., Silva, R., and Fonseca, L. C. (2015). Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão. In *Anais do SBIE*, volume 26, page 1187.
- Silva Filho, R. L. L., Motejunas, P. R., Hipólito, O., and Lobo, M. (2007). A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, 37(132):641–659.
- Sobrinho, J. D. (2008). Avaliação educativa: produção de sentidos com valor de formação. *Revista da Avaliação do Ensino Superior, Campinas*, 13(1):193–207.