# Utilização de técnicas de Mineração de Dados Educacionais para a predição de desempenho de alunos de EaD em Ambientes Virtuais de Aprendizagem

**Humberto Rabelo[2], Aquiles Medeiros Filgueira Burlamaqui[3], Ricardo Alexsandro de Medeiros Valentim[4], Danieli Silva de Souza Rabelo[4], Soraya Roberta dos Santos Medeiros[2]**

[1]Programa de Pós-Graduação em Engenharia Elétrica e Computação
Universidade Federal do Rio Grande do Norte (UFRN)
59078-970 – Natal – RN – Brasil

[2]Departmento de Computação e Tecnologia – Centro de Ensino Superior do Seridó
Universidade Federal do Rio Grande do Norte (UFRN) – Caicó – RN – Brasil

[3]Escola de Ciências e Tecnologia
Universidade Federal do Rio Grande do Norte (UFRN) – Natal – RN – Brasil

[4]Secretaria de Educação à Distância (SEDIS)
Universidade Federal do Rio Grande do Norte (UFRN) – Natal – RN – Brasil

{hrabeloufrn, aquilesburlamaqui, ricardo.lahb, rabrlodanni, soraya.roberta.js}@gmail.com

***Abstract.*** *This work reports application of data mining techniques in an virtual learning environment (AVA) that uses Moodle like distance learning courses platform. Educational data mining produces methods and techniques that seek patterns discovery for providing usable insights on prediction of teaching and learning processes. Experiment uses real data from a historical basis, which contains thirteen classes of undergraduate courses, being part of a study that seeks to improve distance learning process of Federal University of Rio Grande do Norte.*

***Resumo.*** *Este trabalho relata a aplicação de técnicas de mineração de dados em ambiente virtual de aprendizagem (AVA) que utiliza o Moodle como plataforma para realização de cursos de graduação à distância. A mineração de dados educacionais produz métodos e técnicas que objetivam a descoberta de padrões que forneçam conhecimentos utilizáveis na predição dos processos de ensino e aprendizagem. O experimento utiliza dados reais de uma base histórica contendo treze turmas de cursos de graduação, sendo parte de um estudo que visa melhorar o processo de ensino à distância da Universidade Federal do Rio Grande do Norte (UFRN).*

## 1. Introdução

Existem diversas modalidades de Ensino na área da Educação, sendo uma delas a Educação à Distância (EaD), que se caracteriza por sua prática pedagógica diferenciada no processo de ensino e aprendizagem. Tal modalidade de ensino acontece por meio da utilização de tecnologias da informação e comunicação em seus mais distintos graus

DOI: 10.5753/cbie.sbie.2017.1527 

educacionais. Dessa forma, é em cenários como esse que ocorrem diferentes interações entre os alunos e o AVA, o que contribui para a geração de um valor expressivo de dados, os quais, ao serem gerenciados e analisados, expressam limitações ou podem sugerir ampliações sobre os agentes e sua dinâmica de interação com o sistema.

Segundo [Romero & Ventura 2013], gerenciar dados é um dos maiores desafios enfrentados pelas instituições educacionais, uma vez que a quantidade deles tem crescido de maneira exponencial. Assim sendo, a motivação que permeia o problema em questão consiste no fato do ambiente escolar ter se tornado imerso de elementos tecnológicos, tais como *e-learning*, softwares e ambientes educacionais, dentre outros, contribuindo para este aumento de dados.

Diante de tal cenário, tem crescido nos últimos anos a intenção em utilizar a mineração de dados em campos distintos da área científica [Baker, 2011],[Sweeney et al. 2016], inclusive no processo de coleta, análise, transformação e interpretação de dados voltados para a educação, em que se destacam técnicas associadas à Mineração de Dados Educacionais (MDE), do inglês "*Educational Data mining*" (EDM). É por meio dela que se possibilita a criação e a consequente utilização de modelos para descobrir padrões e novas informações sobre os conjuntos de dados acerca dos ambientes de aprendizagem, seus sujeitos e as suas configurações.

Isto posto, o presente estudo objetiva descrever a aplicação de técnicas de Mineração de Dados Educacionais por meio de árvores de decisão, através de um ambiente virtual de aprendizagem educacional, a plataforma Moodle, durante a realização de cursos de graduação à distância pela UFRN. Com esse modelo, pretende-se predizer se o aluno terá sucesso ou insucesso acadêmico no decorrer do curso, mediante seu nível de participação, interação ou desempenho na plataforma.

Tomando como base os pensamentos de [Hansen et al.,1996], [Frield; Brodley, 1997], optou-se por trabalhar com árvores de decisão por sua classificação ser mais objetiva, prover fácil interpretação, bem como ser computacionalmente eficiente. Além disso, [Chikalov, 2011], discorre que as árvores oferecem meios que direcionam ao conhecimento proposicional, a fim de auxiliar no processo de tomada de decisão e classificação preditiva de objetos, tais como o desempenho do aluno no AVA.

O presente trabalho está organizado em cinco seções. A seção dois apresenta a importância da mineração de dados e trabalhos relacionados. A terceira seção discorre sobre a metodologia aplicada. A quarta seção relata os resultados obtidos e análise desses, seguida de considerações finais.

## 2. Mineração de Dados e Trabalhos relacionados

Desde o surgimento dos sistemas computacionais, um dos principais objetivos das organizações tem sido o de armazenar dados. Nas últimas décadas, essa tendência ficou ainda mais evidente com a queda nos custos para aquisição de hardware, tornando possível armazenar quantidades cada vez maiores de dados. Novas e mais complexas estruturas de armazenamento foram desenvolvidas, como, por exemplo, *Data Warehouses*, Bibliotecas Virtuais etc. Observando-se que as técnicas tradicionais de exploração de dados não são mais adequadas para tratar a grande maioria dos repositórios e considerando que o volume de dados armazenados cresce permanentemente, tornou-se difícil responder à seguinte pergunta: o que fazer com os dados armazenados? Com a finalidade de responder a essa questão foi proposta, no final

da década de 80, a Mineração de Dados, do inglês " *Data Mining*" (DM).

A mineração de dados é uma das tecnologias mais promissoras da atualidade, por permitir a coleta e o armazenamento de uma grande quantidade de informações que, ao serem analisadas, podem se tornar conhecimentos de grande valor para as instituições, bem como para a comunidade científica. De acordo com [Witten 2005], existem algumas áreas nas quais a mineração de dados é aplicada satisfatoriamente, tais como na área bancária, na Medicina e no processo de tomada de decisão.

No campo educacional e com o crescente uso dos AVA na *web* e outras tecnologias para apoio ao processo de ensino e aprendizagem, um grande volume de dados tem sido gerado a partir das diferentes modalidades de interação no sistema, envolvendo principalmente estudantes, tutores e professores. Entretanto, boa parte desses dados não têm sido analisados, o que se constitui uma lacuna importante, dada a quantidade de informação valiosa que se pode potencialmente extrair de tais dados.

De acordo com [Gottardo 2014], em busca de melhor compreensão do comportamento dos estudantes e professores e da forma como eles interagem no ambiente virtual, o trabalho realizado por pesquisadores em MDE tem investido no uso e na melhoria de técnicas conhecidas para obter conhecimentos relevantes a partir desses dados. As pesquisas de [Santos et al. 2012], por exemplo, relatam um estudo de caso que permite identificar alunos com maior risco de reprovação. Os modelos criados identificam a probabilidade de reprovação com estimativa de acerto em 69%. Por meio de sua pesquisa, os autores comprovaram que quanto maior a dedicação em atividades presenciais e semipresenciais, melhor o desempenho do aluno.

Em outro campo, o trabalho de [Da Costa et al. 2014] teve como objetivo a identificação do perfil de alunos propensos à evasão. Para tanto, foram utilizados os dados disponíveis em planilhas de cursos de especialização à distância. Os primeiros dados pré-processados e transformados foram classificados utilizando o algoritmo J48 que obteve 97,6% de precisão. Assim sendo, [Moore and Kearsley 2007] abordam que tais conhecimentos podem servir de subsídio para a melhoria das práticas em um contexto de EaD, além de ser ferramenta importante para viabilizar a personalização do ensino.

Nesta perspectiva, este artigo se diferencia dos demais por propor uma nova forma de seleção dos indicadores, direcionada por meio da estatística descritiva. Tal proposta elencou como mais significativos a combinação de uma ação com um módulo, e não somente a exibição da ação como é comumente abordado em vários trabalhos.

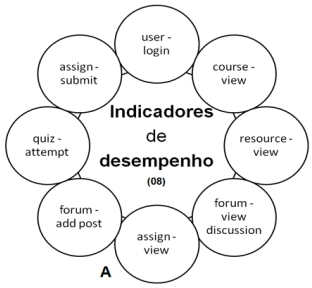
## 3. Metodologia

Inicialmente, tal como [Camilo and Silva 2009], foi realizada uma compreensão do funcionamento do ambiente Moodle e do seu repositório de dados para torná-los representativos, seguida de uma etapa de preparação (filtragem e refinamento) dos dados. Tomando como base as interações armazenadas no repositório do ambiente de pesquisa, utilizando SQL, Linguagem de Consulta Estruturada, com o aplicativo pgadmin3, foram encontradas 64 ações associadas a 24 módulos (atributos action e module, encontrados nas tabelas de log do Moodle), realizadas por 514 usuários com perfil de aluno em 13 Turmas de Cursos Graduação da UFRN, perfazendo um montante de 11.310.848 interações armazenadas no log do Moodle.

A partir dos dados obtidos, realizou-se uma etapa de refinamento desses por

meio de estatística descritiva, média, desvio padrão e coeficiente de correlação de Pearson [Silvestre 2007], em que foram selecionadas dentre as 64 ações existentes aquelas consideradas mais representativas para compor os indicadores adotados e fazer a predição do desempenho dos alunos quanto à utilização do ambiente virtual nos cursos de graduação EaD da Universidade.

Os indicadores de desempenho selecionados para esse experimento foram: (i) Ação login de usuário (login); (ii) Ação de visualização do Curso (view); (iii) Ação de visualização de Recursos do Curso (view); (iv) Ação de Visualização de discussão em fórum (view discussion); (v) Ação de Adição de postagem em fórum (add post); (vi) Ação de visualização de tarefa – View;(vii) Ação de enviar tarefa (Submit) e (viii) Ação de responder questionário (Attempt). São apresentados na Figura 1 os valores dos atributos module e action utilizados para seleção dos atributos para a definição dos índices de desempenho do aluno.




A

| Módulo | Module | Action(s) |
|---|---|---|
| Usuário | User | Login |
| Curso | Course | View |
| Recursos | Resource | View |
| Fórum | Fórum | View discussion, Add post |
| Tarefa | Assign | View,Submit |
| Questionário | Quiz | Attempt |

B

**Figura 1. Indicadores de Desempenho e Atributos Utilizados. Fonte: A: arquivo pessoal dos autores. B: Tabelas do Moodle.**

Uma vez que os dados foram refinados, partiu-se a transformação dos valores dos dados brutos de todos os indicadores selecionados em dados padronizados. Com base nesses, a classificação de desempenho do aluno no ambiente é efetuada, podendo apresentar três possibilidades: as inseridas no intervalo entre 0 e 1 indicam desempenho regular; entre 1 e 2, desempenho bom; e, entre 2 e 3, desempenho ótimo.

Para aplicar as Técnicas de Mineração de Dados, fez-se o emprego da ferramenta computacional Weka, concentrando-se na seara da Classificação, tomando como referência a taxonomia estipulada por [Baker et al. 2011]. Segundo [Romero et al. 2008], a classificação admite que se possa prever o desempenho do aluno e sua consequente avaliação final, bem como apontar quais discentes encontram-se desmotivados durante o curso em que estão matriculados, o que vem a corroborar com a proposta objetivada no presente trabalho. Conforme mencionado em seção anterior, a técnica de classificação empregada neste trabalho é a de "árvore de decisão", dada a sua eficiência computacional e representação de informação, a qual se mostra adequada para o problema analisado neste artigo.

Seguindo esses pressupostos, utilizou-se dois algoritmos de classificação baseados em árvores de decisão, a saber, ID3 e J48. O Algoritmo ID3 consiste em construir uma árvore de decisão com base em um processo de análise dos atributos do conjunto de dados apresentado. Outro algoritmo utilizado é o J48, o qual emprega a estratégia dividir para conquistar. Uma das motivações ao utilizar o J48 neste trabalho é o fato de ele desempenhar relevante contribuição sobre as variáveis qualitativas

presentes nas bases de dados utilizadas pelo presente trabalho, construindo e classificando árvores de decisão que apresentam em suas ramificações os atributos de maior relevância. Apresenta-se, a seguir, a árvore gerada pelo Algoritmo J48 da ferramenta Weka:

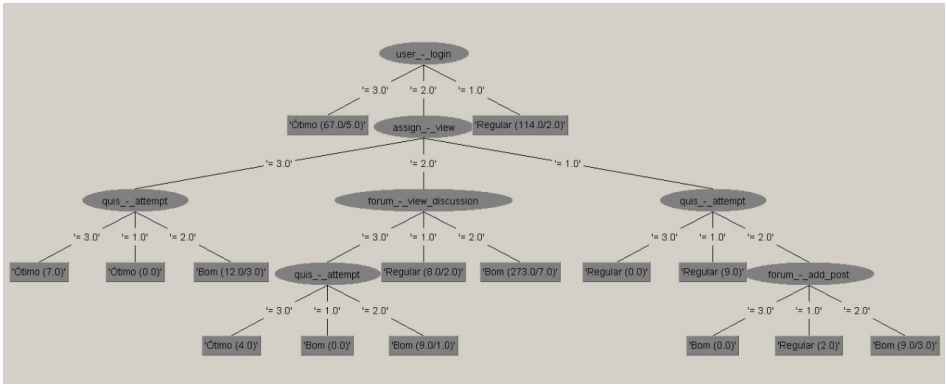



**Figura 2. Árvore de decisão. Fonte: arquivo pessoal dos autores.**

## 4. Resultados e Discussão

É possível observar na Tabela 1, que o experimento foi realizado com 514 instâncias de alunos e que foram encontradas 483 instâncias classificadas corretamente, o que corresponde a 93,97% de acertos obtidos através do classificador ID3, e 496 instâncias classificadas corretamente, correspondendo a 96,50% de acertos através do classificador J48. Sendo encontradas 31 instâncias classificadas incorretamente, o que corresponde a 6,03% de erros obtidos através do classificador ID3, e 18 instâncias classificadas incorretamente, correspondendo a 3,50% de erros através do classificador J48. O melhor caso foi obtido através do algoritmo J48, com 96,5% de acertos e 3,5% de erros de classificação. Desse modo, o pior caso foi com algoritmo ID3 com 93,97% de acertos e 6,03% de erros de classificação com o mesmo modelo preditivo com os mesmos indicadores:

**Tabela 1. Acurácia dos classificadores. Fonte: arquivo pessoal dos autores.**

| *DataSet* | ID - 3 | | | | J-48 | | | |
|---|---|---|---|---|---|---|---|---|
| | instâncias Classificadas Corretamente | | instâncias Classificadas Incorretamente | | instâncias Classificadas Corretamente | | instâncias Classificadas | |
| **Instâncias (alunos)** | Valor | Percentual | Valor | Percentual | Valor | Percentual | Valor | Percentual |
| **514** | **483** | 93,97 | **31** | 6,03 | **496** | 96,50 | **18** | 3,50 |
| *Accuracy* | 0.939 | | | | 0.965 | | | |
| *Kappa statistic* | 0.8931 | | | | 0.9385 | | | |

O gráfico da Figura 3 exibe o resultado comparativo em percentual das classificações das instâncias consideradas como corretas e incorretas utilizando o algoritmos ID3 e J48 no experimento realizado.

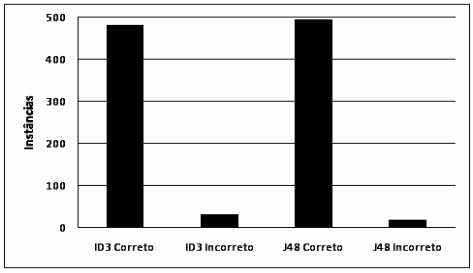



**Figura 3. Comparativo dos classificadores. Fonte: Arquivo pessoal dos autores.**

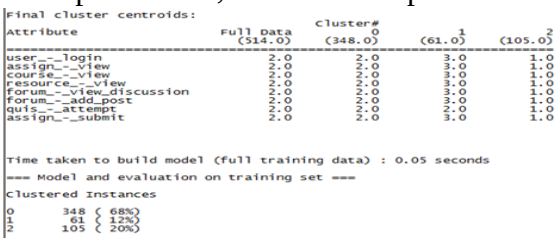
Tão importante quanto o número de instâncias corretamente classificadas é a estatística de Kappa, métrica que indica o nível de concordância (coesão) dos dados classificados. Nesse experimento, o valor do índice Kappa variou de 0,89 a 0,93, considerado excelente, uma vez que sua variação vai de 0 a 1 de magnitude. Na Tabela 4, é possível observar a distribuição dos Alunos das Classes, (Regular, Bom e Ótimo) obtidas por meio de Estatística descritiva e normalização.

**Tabela 2. Distribuição das classes obtidas. Fonte: arquivo pessoal dos autores.**

| Nome da Classe | Descrição | Número de Alunos | Percentual de Alunos |
|---|---|---|---|
| **Regular** | Alunos com desempenho abaixo da média | 135 | 26,26 % |
| **Bom** | Alunos com desempenho mediano. | 298 | 57,98 % |
| **Ótimo** | Alunos com desempenho acima da média | 81 | 15,76 % |
| **Total** | Alunos do experimento | 514 | 100 % |

A Clusterização, é o processo de agrupamento de um conjunto de objetos dentro de classes de objetos similares [Jiawei and Han 2005]. Para a execução deste agrupamento do experimento, o atributo classe foi desconsiderado, porém utilizado posteriormente para verificar a qualidade da clusterização obtida. Utilizou-se o algoritmo SimpleKmeans, que é baseado no clássico algoritmo K-means. A ideia desse algoritmo é fornecer uma classificação de informações de acordo com os próprios dados.

Dessa maneira, o algoritmo fornecer uma classificação automática sem a necessidade de uma supervisão humana, ou seja, sem nenhuma pré-classificação existente. Por causa dessa característica, o K-means é considerado como um algoritmo de mineração de dados não supervisionado. [Pimentel 2003, pag. 497] o define como uma técnica que usa o algoritmo de agrupamento de dados por K-médias (K-means clustering). O objetivo deste algoritmo é encontrar a melhor divisão de P dados em K grupos Ci, i = 1, ... K, de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada, tal como afirma [Pimentel et. al., 2003]. Para a execução desse algoritmo foi-se utilizado k = 3 e Distância Euclidiana como parâmetros, os resultados podem ser vistos na Figura 4.



```
Final cluster centroids:
                                Cluster#
Attribute               Full Data       0          1          2
                          (514.0)   (348.0)    (61.0)   (105.0)
=============================================================
user_-_login                 2.0       2.0        3.0        1.0
assign_-_view                2.0       2.0        3.0        1.0
course_-_view                2.0       2.0        3.0        1.0
resource_-_view              2.0       2.0        3.0        1.0
forum_-_view_discussion      2.0       2.0        3.0        1.0
forum_-_add_post             2.0       2.0        3.0        1.0
quis_-_attempt               2.0       2.0        2.0        1.0
assign_-_submit              2.0       2.0        3.0        1.0


Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      348 ( 68%)
1       61 ( 12%)
2      105 ( 20%)
```

**Figura 3. Comparativo dos classificadores. Fonte: Arquivo pessoal dos autores.**

Por meio da Figura 4, pode-se observar que alguns atributos não agrupam bem nos clusters do SimpleKmeans, por exemplo, o atributo quis-attemp que teve seu valor 2 (Classe Bom Desempenho) pertencente tanto ao cluster 0 como ao 1. Algumas das instâncias podem não ser clusterizadas corretamente. A fim de obter uma possível melhoria na clusterização ou agrupamento das classes, e também para comparação, optou-se por classificar aos alunos nas Classes definidas (Regular, Bom e Ótimo) por meio de estatística descritiva, uma vez que se dispõe dos quantitativos dos dados.

### 4.1. Matriz de confusão

Uma matriz de confusão contém informações importantes para o entendimento do resultado do algoritmo, dentre elas: a quantidade de instâncias classificadas corretamente; a quantidade de instâncias classificadas erroneamente; a quantidade de instâncias que o algoritmo acreditava ser de um tipo (Bom) e na verdade foram classificadas como (Regular) ou com (ótimo) por exemplo. A matriz de confusão destacada na Tabela 3, fornecendo informações adicionais sobre os resultados obtidos na execução do algoritmo J48.

**Tabela 3. Matriz de Confusão Algoritmos - Fonte: arquivo pessoal dos autores.**

| J-48 | | CLASSES PREVISTAS | | |
|---|---|---|---|---|
| | | Ótimo | Bom | Regular |
| CLASSSES CORRETAS | Ótimo | 75 | 6 | 0 |
| | Bom | 5 | 289 | 4 |
| | Regular | 0 | 3 | 132 |

| ID-3 | | CLASSES PREVISTAS | | |
|---|---|---|---|---|
| | | Ótimo | Bom | Regular |
| CLASSSES CORRETAS | Ótimo | 67 | 14 | 0 |
| | Bom | 6 | 286 | 6 |
| | Regular | 0 | 5 | 130 |

O algoritmo J48 obteve acurácia de 96,5%, classificando corretamente o desempenho de 496 dos 514 alunos. A matriz de confusão da Tabela 3 fornece informações adicionais sobre os resultados do experimento. É possível observar que seis instâncias foram classificadas como pertencendo à classe Bom, quando de fato pertencem à classe Ótimo. Da mesma forma, cinco instâncias foram classificadas incorretamente como Ótimas, e quatro como Regular quando na verdade ambas pertencem a classe Bom, e por fim temos outras três instâncias classificadas como Bom, quando de fato pertencem a classe Regular.

Quanto aos resultados obtidos na execução do algoritmo ID3 que obteve acurácia de 93,9%, classificando corretamente o desempenho de 483 dos 514 alunos. A partir da matriz de confusão da Tabela 3, é possível observar que 14 instâncias foram classificadas como pertencendo a Classe Bom, quando de fato pertencem a classe Ótimo, e da mesma forma 06 instâncias foram classificadas como Ótimas, e 6 como Regular quando na verdade pertencem a classe Bom, e por fim outras 05 instâncias foram classificadas como Bom, quando de fato pertencem a classe Regular.

### 4.2. Validação da Hipótese

A aprovação dos alunos em uma disciplina é fundamental para validar hipóteses de problema e de solução, de que o aluno com desempenho efetivo de participação no AVA apresentará nota final melhor e, consequentemente, melhor taxa de aprovação na disciplina, como foi considerado na construção do Modelo Preditivo. Entretanto, nem todas as turmas possuem as notas finais lançadas no ambiente virtual, uma vez que ele é considerado um apoio para as disciplinas de graduação em EaD da instituição, e que a obrigatoriedade de lançamento das notas dos alunos ocorre no Sistema Integrado de Gestão de Atividades Acadêmicas. Dessa forma, para validar esta hipótese, será analisada uma das turmas que tem 268 alunos e suas notas finais disponíveis no Ambiente.

Ao final do procedimento de mineração de dados realizado nessa turma, fez-se a predição de que, dos 268 alunos, 64 seriam classificados com desempenho Ótimo, e 164 com desempenho Bom, e 40 com Regular. Segundo os dados reais desses mesmos alunos, 234 foram aprovados e 34 foram reprovados. Em linhas gerais, a predição de aprovados foi de 228, quando se deveria encontrar 234. Para os reprovados, a predição

foi de 40, quando se deveria encontrar 34, ocasionando em uma diferença de seis alunos, entre o previsto e o real.

**Tabela 4. Sumarização dos resultados. Fonte: Arquivo pessoal dos autores.**

| | Resultados | Correto | Incorreto |
|---|---|---|---|
| **Aprovado** | **234** | **210** | **24** |
| Bom | 149 | 149 | - |
| Ótimo | 61 | 61 | - |
| Regular | 24 | - | 24 |
| **Reprovado** | **34** | **16** | **18** |
| Bom | 15 | - | 15 |
| Ótimo | 3 | - | 3 |
| Regular | 16 | 16 | - |
| **Total de alunos** | **268** | **268** | **268** |

Ao sumarizar os resultados, pela Tabela 4, temos de fato dos 234 alunos realmente aprovados da turma, 210 alunos foram classificados corretamente como aprovados, e 24 alunos foram classificados incorretamente, pois tiveram desempenho regular e foram aprovados, quando o esperado é que fossem reprovados. Quanto aos 34 alunos realmente reprovados da turma, 16 alunos foram classificados corretamente como reprovados, enquanto 18 alunos, sendo destes 15 classificados com desempenho bom, e 3 classificados com desempenho ótimo que foram previstos como reprovados, quando o esperado é que fossem aprovados.

Para testar a Hipótese utilizou-se o teste de Qui-Quadrado, que objetiva verificar se a frequência absoluta observada de uma variável é significativamente diferente da distribuição de frequência absoluta esperada. Temos as duas hipóteses abaixo:

H0: O resultado final de um aluno (aprovação) é independente do seu desempenho (ou nível de participação no ambiente virtual;

H1: O desempenho do aluno no ambiente virtual, influencia em sua aprovação na disciplina;

μ = 0,05 (significância do teste) g.l = onde consultando na tabela padrão, temos que Qui-quadrado tabelado é igual a 3,84.

Calculando o Qui-quadrado, com dados concretos e dados esperados temos:

X2 = (((228-231)2)/231)+(((234-231)2)+(((40-37)2)/37)+(((34-37)2)/37)) = 9,52

Como Qui-Quadrado calculado (9,52) é maior que Qui-quadrado tabelado (3,94) com 1g.l. ao nível de significância de 5%, rejeita-se H0 em prol de H1. Conclui-se, portanto, que há evidências de que o desempenho ou participação efetiva do aluno no ambiente virtual influencia no resultado final ou aprovação do aluno.

## 5. Considerações finais

Neste trabalho, foi possível notar a relevância do presente estudo da área de MDE aplicado em dados reais para a tomada de decisão de uma Instituição de Ensino, bem como seus docentes, posto que através da técnica de mineração de dados que utiliza algoritmos de Classificação por árvores de decisão a acurácia ficou entre 93,9% e 96,5% de precisão se um aluno terá ou não um desempenho satisfatório, não sendo necessário esperar pelo final da disciplina para saber o desempenho final desses discentes, contribuindo para propor ações de ajuste de conduta durante o processo de ensino e aprendizagem e colaborando para diminuir os índices de evasão.Ademais, após a realização de todas as etapas do experimento, pode-se constatar que a definição ou a escolha dos atributos de entrada, para criar os indicadores utilizados nas estratégias é um

fator determinante para o sucesso destas, bem com os parâmetros de configuração dos algoritmos.

Por fim, para trabalhos futuros, sugere-se que sejam feitos estudos detalhados e testes nas configurações de algoritmos de outros tipos de técnicas de mineração de dados com a finalidade de melhorar ainda mais o desempenho e precisão dos resultados, e propõe-se ainda automatizar o processo de seleção dos indicadores de desempenho, através de implementação de uma meta-heurística onde o usuário seja capaz de selecionar os atributos action e module desejados, que possam representar o modelos preditivos de conhecimento de algum aspecto que ele deseje acompanhar, verificar ou prever, de forma que possibilite se adaptar a cada necessidade.

## Referências

Baker, R., Isotani, S., & Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. Brazilian Journal of Computers in Education, v.19, n. 02, p. 03, 2011.

Baker, R. S. J. D. (2010). Data mining for education. International encyclopedia of education, 7(3), 112-118.

Camilo, C. O., & Silva, J. C. D. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFC), 1-29.

Chikalov, I. (2011). Average Time Complexity of Decision Trees (Vol. 21). Springer Science & Business Media.

Hansen, M.; Dubayah, R.; Defries, R. Classification trees: an alternative to traditional land cover classifiers.International Journal of Remote Sensing , v. 17, n. 5, p. 1075-1081, 1996.

Da Costa, s. S., Cazella, S., & Rigo, S. J. (2014). Minerando Dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EaD: Um estudo de caso sobre evasão escolar na UNA-SUS. RENOTE, 12 (2)

Fontana, A., & Naldi, M. C. (2009, March). Estudo e comparação de métodos para estimação de números de grupos em problemas de agrupamento de dados. ICMC.

Gottardo, Ernani; Kaestner, Celso Antônio Alves; Noronha, Robinson Vida (2014). Estimativa de Desempenho Acadêmico de Estudantes: Análise da Aplicação de Técnicas de Mineração de Dados em Cursos a Distância. Revista Brasileira de Informática na Educação, [s.l.], v. 22, n. 01, p.45-55, 18 maio 2014.

Han, J. (2005). Data Mining: Concepts and Techniques. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Kearsley, G., & Moore, M. (2007). Educação a Distância: uma visão integrada. São Paulo: Thomson Learning.

Pimentel, E. P, França, V. F., and Omar, N. (2003). A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In Anais do Simpósio Brasileiro de Informática na Educação,Rio de Janeiro, RJ. SBC. p.495-504.

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12-27.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management

systems: Moodle case study and tutorial. Computers & Education, 51(1), 368-384.

Santos, H., Camargo, F., & Camargo, S. (2012). Minerando Dados de Ambientes Virtuais de Aprendizagem para predição de Desempenho de Estudantes. Latin American Conference on Learnig Objects and Technologies (LACLO), 3 (1).

Sweeney, Mack et al. Next-Term Student Performance Prediction: A Recommender Systems Approach. Jedm - Journal Of Educational Data Mining. Memphis, Tennessee, p. 22-51. jul. 2016. Disponível em: <https://jedm.educationaldatamining.org/index.php/JEDM/index>. Acesso em: 02 jul. 2016.

Silvestre, A. (2007). Análise de dados e estatística descritiva. Escolar editora.

Weka - Waikato Environment for Knowledge Analysis. Weka 3: Data Mining Software in Java. Disponivel em <http://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 13/04/2016.

Witten, I. H., & Frank E. (2005). Data Mining: Practical machine learning tools and techniques, 2, 127-143.