

Análise de perfis de interação de alunos no ambiente virtual Moodle via agrupamento hierárquico

**Pryscilla de Sousa Pereira¹, Joana Laysa Lima Cunha¹,
Gilvandenys Leite Sales¹,
Ananda Lima Freire²**

¹Programa de Pós-Graduação em Ciência da Computação – PPGCC
Instituto Federal de Educação, Ciência e Tecnologia do Ceará – IFCE
Av. Treze de Maio, 2081 – Benfica – 60.040-215 – Fortaleza – CE – Brazil

²Instituto Atlântico
Fortaleza – CE – Brazil

{pry.spereira,laysa.cunha, denyssales}@gmail.com, ananda-freire@atlantico.com.br

Abstract. *This article presents a student's activities monitoring approach using Data Mining, with the objective of identifying students groups with the same profile following some characteristics, like participation and interaction on LMS by applying a hierarchical clustering. This proposal is based on Learning Vectors Model. The results show 3 distinct groups (great, good and regular) taking into consideration the interactions of the environment. So allowing the recognition of difficulties, and making easy for tutors to act directly on these difficulties and so improve the performance and learning from the students.*

Resumo. *Este artigo apresenta uma abordagem de monitoramento das atividades dos estudantes, usando técnicas de Mineração de Dados. O objetivo deste estudo é identificar grupos de alunos com o mesmo perfil de participação e interação em um ambiente de aprendizagem on-line, através de algoritmo de agrupamento hierárquico. Esta proposta baseia-se no modelo de avaliação Learning Vectors e usa como fonte de informação os logs retirados do banco de dados Moodle. Os resultados apresentaram 3 grupos distintos (ótimo, bom e regular) em relação a interação no ambiente. O que permitiu o reconhecimento de suas dificuldades, facilitando o direcionamento das intervenções do tutor para melhorar a performance e a aprendizagem do aluno.*

1. Introdução

A educação no Brasil está trilhando um novo caminho com o crescimento da Educação a Distância (EaD), proporcionando uma formação de qualidade a qualquer pessoa, independente da sua disponibilidade e localidade [ABED 2017]. A EaD tem como objetivo oferecer um processo de aprendizagem completo, dinâmico e eficiente por intermédio de recursos tecnológicos e, para dar suporte a essa modalidade, existem os Ambientes Virtuais de Aprendizagem (AVA). Esses ambientes são a sala de aula da EaD, é o meio pelo qual professor, tutor e aluno se comunicam, sem a necessidade de estarem no mesmo local geográfico e no mesmo horário.

Os AVA's podem ser definidos como um sistema computacional implementado por meio de uma linguagem de programação que reúne possibilidades de acesso on-line ao conteúdo de cursos, além de diversos recursos de comunicação e interação entre

os sujeitos que participam do ambiente, em um único software denominado plataforma [Passerino and Gerling 2013].

Várias pesquisas estão voltadas para a melhoria dos AVA's, tentando tornar a sala de aula virtual mais próxima da presencial, diminuindo a distância transacional entre os envolvidos e buscando promover a interatividade de diversas formas. A teoria da distância transacional sugere que a Educação a Distância traz uma separação psicológica e comunicacional existente entre professores e alunos quando distanciados no espaço e no tempo. Essa separação gera possíveis incompreensões que precisam ser diminuídas para que a aprendizagem possa ocorrer de modo efetivo [Moore 2013].

O tutor tem papel fundamental na abordagem pedagógica para diminuir essa distância, destacando em suas atribuições a rápida correção dos exercícios propostos, para que o mesmo tenha a chance de interferir na aprendizagem e fazer o acompanhamento necessário [Leite et al. 2015]. No entanto, nem sempre consegue se fazer tão presente nesse acompanhamento, pois existem turmas numerosas e que demandam muito tempo nas correções de atividades e interações no AVA. Essa demanda, faz com que o tutor não perceba as dificuldades dos alunos e sua falta de motivação, acarretando em diminuição do rendimento e possíveis evasões [dos Santos and Falcão 2017].

Com o intuito de diminuir os problemas enfrentados pelos tutores, pode-se destacar o modelo *Learning Vectors* (LV), implementado em cursos de EaD. Esse modelo tem sua metodologia fundamentada na interação e auxilia no acompanhamento da avaliação das atividades síncronas e assíncronas, através de uma avaliação qualitativa baseada em ícones associados aos conceitos: Muito Bom, Bom, Regular, Fraco, Não Satisfatório e Neutro [Sales 2010].

Além do Modelo LV, outros estudos estão sendo desenvolvidos, como os trabalhos de [Akçapınar et al. 2014, Młynarska et al. 2016, Gottardo et al. 2012] que utilizam Técnicas de Inteligência Artificial e Mineração de Dados integrados ao AVA permitindo explorar os dados gerados com o objetivo de facilitar ao máximo o trabalho dos professores e tutores, permitindo que eles se preocupem apenas com a aprendizagem do aluno. [Baker et al. 2011] discursa sobre a importância da Mineração de Dados Educacionais para o Brasil e a oportunidade de melhorar a educação do país. Ele enfatiza que, com a utilização da coleta de dados em grande escala, é possível criar modelos e fazer previsões aplicáveis a qualquer AVA, inclusive em salas de aula convencionais.

Nesse contexto, este trabalho pretende analisar os dados gerados em AVA com o intuito de identificar padrões de comportamento dos alunos, agrupando-os em níveis de risco de reprovação e/ou evasão e de baixo nível de aprendizagem, para que o tutor possa fazer as intervenções necessárias com os alunos identificados. A partir dessas intervenções, espera-se que os alunos saiam dos grupos de risco e possam ter um outro comportamento dentro do ambiente, assim melhorando sua aprendizagem e, consequentemente, suas notas.

Este artigo está organizado com a seguinte estrutura: a próxima seção apresenta o Modelo LV, assim como os trabalhos relacionados à Mineração de Dados no acompanhamento de estudantes. A seção 3 expõe a metodologia de desenvolvimento deste trabalho. Na seção 4 é realizada a apresentação e discussão dos resultados e na seção posterior a apresentação das conclusões desta pesquisa.

2. Trabalhos Relacionados

Com o avanço da tecnologia, vários estudos acerca do AVA estão sendo desenvolvidos com o objetivo de melhorar o desempenho dos alunos e propor ferramentas que diminuam a sobrecarga dos tutores, além de buscar promover a interatividade de diversas formas.

2.1. Modelo *Learning Vectors*

O Modelo LV [Sales 2010] é um sistema de avaliação que preza pela colaboração e interação entre os atores envolvidos na EaD: Professores/Tutores, Alunos e AVA. Também apresenta métricas alinhadas a um sistema de avaliação quali-quantitativo e permite guiar o processo de avaliação na direção formativa.

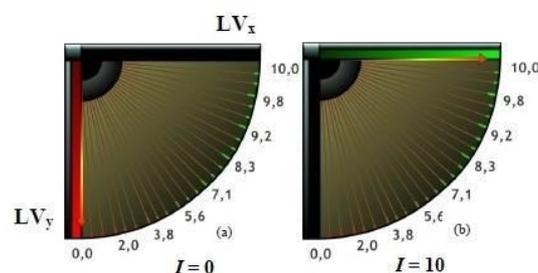


Figura 1. Vetor de Aprendizagem. Fonte: [Sales 2010]

O autor utiliza preceitos matemáticos para medir a quantidade e a qualidade de interações e atividades realizadas pelos alunos, utilizando vetores em duas dimensões, ou melhor, em dois eixos, que representam a Positividade e a Negatividade, chamado de Vetores-Aprendizagem (FIGURA 1). Esse vetor possui módulo fixo de 10 unidades e direção variável, movendo-se em sentido anti-horário, cujos ângulos de avanço relacionam-se a uma menção qualitativa atribuída pelo Tutor.

O Modelo LV também utiliza-se de ícones para atribuir valores, com os quais se propõe transmitir a ideia de afetividade e proximidade. Esses ícones são atribuídos pelo Tutor às ações dos alunos, sejam nas ferramentas de interação e/ou nas atividades postadas, correlacionando-se a valores numéricos de notas. Uma outra métrica associada ao Modelo LV é o Fator β , que é um indicador qualitativo não-linear para suporte no acompanhamento da aprendizagem do aluno, adaptado a partir da dimensão psicológica Positividade/Negatividade (Taxa P/N) do modelo de avaliação de desempenho denominado Meta *Learning* (Modelo ML) [Losada and Heaphy 2004]. Esse fator expressa o desempenho do aluno em: Muito Alto, Alto, Médio, Baixo e Muito Baixo, com valores médios classificados [Sales 2010].

Essa ferramenta possibilita o *feedback* em tempo mínimo e reduz a sobrecarga de trabalho gerada para o Tutor [Sales 2010]. Além disso, o Modelo LV avalia o aprendizado qualitativamente e, também, pode ser aplicado para diferenciar estudantes com notas finais iguais ou próximas através do Fator β e, além de auxiliar o tutor a traçar critérios para julgar o nível de aprendizagem do aluno para então aprová-lo ou reprová-lo ao final de um curso.

2.2. Mineração de Dados Educacionais

O acompanhamento do desempenho de estudantes na Educação a Distância é complexo, subjetivo e gera alta demanda de trabalho para professores e tutores. Desta forma, a Mineração de Dados tem sido utilizada na área educacional com o objetivo de auxiliar docentes.

No trabalho de [Akçapınar et al. 2014] e [Talavera and Gaudioso 2004] propõe-se identificar grupos de alunos que exibissem comportamentos de interação semelhantes em ambientes de aprendizagem on-line. Ambos utilizaram dados referentes as ações dos estudantes, o primeiro diferencia-se por utilizar apenas informações dos fóruns de discussão e de navegação. O autor utilizou o algoritmo *Model-Based Clustering*, obtendo 6 grupos de alunos de acordo com o perfil apresentado, já o segundo trabalho utiliza o algoritmo de agrupamento SOM (*Self-organized map*) gerando 3 grupos de estudantes bem definidos. Em ambos observou-se que os alunos que possuem alta iniciativa, promovem discussões e têm elevada participação no fórum compreendem ao grupo de alunos com maior percentual de sucesso no curso.

[Gotardo et al. 2013] e [Kampff et al. 2008], por sua vez, aplicaram métodos de aprendizagem supervisionada de forma a apoiar a tomada de ações pró-ativas, chamando a atenção do docente sobre situações que podem levar à evasão ou reprovação. Ambos utilizaram vários dados dos alunos gerados a partir de um AVA. O segundo trabalho obteve como resultado que a quantidade de acessos aos materiais de estudo, o volume de postagens no fórum e a realização de todas as atividades influenciam em resultados acima da média. Já o primeiro autor obteve inferências sobre a aprovação com taxas de precisão próximas a 75%, mesmo em períodos iniciais do curso.

Na pesquisa de [Młynarska et al. 2016] 3 hipóteses são trabalhadas: a submissão antecipada melhora os resultados, um alto nível de atividade é preditivo de bom desempenho e a atividade noturna é ainda melhor do que a atividade diurna. Para validar as hipóteses aplicou-se a correlação de *Pearson*. Para esse experimento, apenas dados relacionados a notas, prazos, horários de submissão e registros de atividades gerais foram considerados. Os experimentos mostraram que os alunos que são mais ativos no AVA e enviam tarefas mais cedo alcançam melhores resultados, no entanto, também mostraram um problema patológico em que altos níveis de atividade se correlacionam com resultados ruins, o que faz necessário novos experimentos.

Os dados usados nos trabalhos citados foram extraídas através de questionários ou logs de acessos de algum AVA, principalmente, o Moodle. Logo, esta pesquisa difere das apresentadas por utilizar o Modelo LV para a extração dos dados e como modelo norteador. Além disso, nos trabalhos citados, as técnicas mais utilizadas são de aprendizagem de máquinas supervisionada, para esta abordagem, é utilizado um algoritmo de agrupamento hierárquico de aprendizagem não-supervisionada.

3. Metodologia

A presente pesquisa foi realizada com dois cursos de pós-graduação da Universidade Federal do Ceará (UFC): um em Educação de Jovens e Adultos e outro em Educação Ambiental. Os dois somam 266 alunos divididos em 10 turmas com uma média de 26 discentes por turma. A disciplina utilizada nos experimentos é denominada Introdução ao Curso

e ao Ambiente Virtual, que é oferecida no primeiro semestre. Nesta, os estudantes são categorizados como aprovados (nota igual ou superior a 7,0), de recuperação ou exame final (nota entre 5 e 6,9) e reprovados (nota abaixo de 4,9). Este curso foi dividido em 3 unidades de aprendizagem, cada uma com várias atividades propostas, tendo como tempo total de duração 41 dias (Unidade 1 - 17 dias; Unidade 2 - 13 dias; Unidade 3 - 11 dias).

O ambiente virtual de aprendizagem utilizado nesse experimento é o Moodle, um dos exemplos de AVA baseado em código livre e aberto [Seixas et al. 2012]. O banco de dados selecionado foi o do ambiente Moodle integrado ao Modelo LV. Todos os registros do sistema, como atividades e interações realizadas pelo professor e aluno durante o curso, são armazenados nesse banco e somam um total de 84192 logs. Todas as atividades propostas neste curso, como Fórum, Wiki e Quiz, seguiram o Modelo LV.

O banco de dados retirado do Moodle foi dividido em três, com os mesmos atributos, porém com períodos diferentes (Unidade 1, Unidade 2 e Unidade 3). Cada período é referente a uma unidade de aprendizagem da disciplina. O primeiro é baseado nas atividades e interações da unidade de aprendizagem 1, o segundo recolheu os dados da unidade de aprendizagem 1 com os dados da unidade de aprendizagem 2 e o terceiro é a união dos dados das unidades de aprendizagem 1, 2 e 3. Consultas SQL foram usadas para extração de dados com a ajuda do software *MySQL Workbench*¹.

No processo de extração de dados, duas bases foram geradas: a primeira referente a dados de visualização retirados a partir de logs de navegação dos alunos e baseado no trabalho de [Akçapınar et al. 2014] e a segunda é mais completa, com atributos retirados de atividades e recursos, que foram baseados no trabalho de [Talavera and Gaudioso 2004] e [Kampff et al. 2008]. Estes atributos são informações numéricas e representam o comportamento de cada estudante.

No entanto, para o atributo tempo de entrega houve uma classificação personalizada, já que os dados disponíveis eram o prazo final, o dia em que a atividade iniciou e a data em que aluno submeteu. Então decidiu-se fazer uma classificação de acordo com a proximidade do limite do prazo. Os alunos foram considerados “muito ágeis” quando entregam dentro de 40 % do tempo máximo indicado, “ágeis” quando entregam dentro de 41 % até 70 % do tempo prazo estipulado e “pouco ágeis” quando entregam dentro de 71 % até 100 % do tempo prazo final. Portanto, é possível verificar quem sempre entrega uma atividade perto do prazo e se a entrega no prazo tem alguma influência no aprendizado. Sendo assim, soma-se um total de 6 bases de dados utilizadas nos experimentos, 3 somente com as visualizações e 3 com os demais atributos selecionados. A Tabela 1 lista todos os atributos que serão utilizados nos experimentos.

O objetivo de utilizar a base de visualizações é verificar se é possível agrupar os alunos de acordo com algum padrão e verificar se esse acesso pode mostrar algo sobre a aprovação dos estudantes. Já com a base contendo todos os demais atributos, além de verificar um padrão, pretende-se analisar quais as atividades que têm mais impacto na aprovação dos alunos e se a entrega adiantada de atividades infere algo sobre o padrão de comportamento dos mesmos.

Após a definição e configuração dos atributos, o próximo passo foi executar o algoritmo de agrupamento. O método de agrupamento hierárquico aglomerativo foi es-

¹Ferramenta de gerenciamento do banco de dados MySQL [McLaughlin 2013]

Tabela 1. Descrição dos atributos selecionados. Fonte: O Autor

	Atributos	Descrição
BASE 2	n_quizLV	número de vezes que realizou o Quiz LV
	p_quizLV	tempo médio para fazer Quiz LV
	n_forumLV	média de posts no Fórum LV
	p_forumLV	tempo médio na postagem no Fórum LV
	c_forumLV	média de comentários por post
	n_tarefasLV	quantidade de tarefas submetidas
	p_tarefaLV	tempo médio para fazer a Tarefa LV
	n_wikiLV	número de edições no Wiki LV
	p_wikiLV	tempo médio para fazer o Wiki LV
BASE 1	a_paginas	número de acessos a páginas
	a_url	número de acessos a url
	a_curso	numero de acessos ao curso
	a_pastas	numero de acessos a pastas disponíveis
	a_quizLV	número de acessos ao Quiz LV
	a_forumLV	número de acessos ao Fórum LV
	a_wikiLV	número de acessos ao Wiki LV
	a_tarefaLV	número de acessos a Tarefa LV
	faltas	percentual de faltas

colhido para ser aplicado. Ele usa a matriz de dissimilaridade que representa o grau de similaridade entre pares de objetos como um dos critérios de agrupamento, ou seja, combina-se amostras com alguma característica comum. O resultado é representado por um dendograma ² [Wilks 2011].

O agrupamento aglomerativo, também conhecido como AGNES (*Agglomerative Nesting*), funciona de forma que cada objeto é inicialmente considerado como um *cluster* de elemento único (folha). Em cada iteração do algoritmo, dois *clusters* muito próximos são combinados em um novo *cluster* maior (nós). Esse procedimento é iterado até que todos os pontos sejam membros de um grupo grande (raiz). Esse algoritmo foi selecionado por trabalhar melhor com o número de amostras da proposta e por analisar as características comuns entre os objetos observados.

O algoritmo AGNES foi implementado usando a linguagem R e o software livre RStudio. Antes de aplicá-lo os dados foram normalizados, subtraindo os valores de sua média e dividindo pelo desvio padrão, para garantir que não exista um atributo dominante. O método foi executado utilizando a métrica euclidiana, que calcula as diferenças entre as amostras, e o método Ward, que mede a dissimilaridade entre os objetos. Como resultado desta execução obtém-se um dendograma, que pode ser cortado para observar os subgrupos. Após algumas análises e a aplicação do método Silhueta para verificar a quantidade ótima de *clusters*, foram selecionados três subgrupos para verificar algum padrão de comportamento entre os alunos.

²Árvore de *cluster*, muito utilizado por algoritmos hierárquicos para visualizar os grupos gerados.

4. Resultados

Nesta seção, são descritos os resultados obtidos a partir dos vários testes realizados. Para melhor entender como se deu o agrupamento, optou-se por utilizar métricas estatísticas para analisar os grupos. Além disso, observou um comportamento que se repetia levando em consideração o perfil de visualização e participação no ambiente dos grupos gerados. Assim, definiu-se por identificar cada cluster como ótimo, bom e regular.

No agrupamento da Unidade 1 - Base 1, percebe-se que o número de visualizações é baixo, levando em consideração a média dos grupos e sabendo que essa unidade durou 17 dias, espera-se que os alunos acessem diariamente o ambiente. O *cluster* com maior número de visualizações é superior ao esperado, no entanto o número de integrantes é tão pequeno que não é possível fazer inferências concisas (TABELA 2). Considera-se que por ser a primeira aula os estudantes ainda estão em processo de adaptação.

Na Unidade 3 - Base 1, o agrupamento mostrou praticamente apenas dois grupos, isso porque o grupo dos alunos ótimos representa um tipo de estudante que não condiz com os demais, eles tem um nível de acesso e participação muito acima em relação aos demais.

Tabela 2. Distribuição de alunos nos grupos gerados. Fonte: O Autor

	Base 1			Base 2		
	Unidade 1	Unidade 2	Unidade 3	Unidade 1	Unidade 2	Unidade 3
Ótimo	18	75	5	97	46	134
Bom	114	135	197	85	158	103
Regular	134	56	64	84	62	29

O agrupamento da Unidade 1 - Base 2, apresenta uma distribuição bem equivalente entre os grupos. Na unidade 1, tem-se que os alunos do grupo ótimo são os mais ativos no ambiente e com o percentual de faltas mais baixo. O quantitativo de alunos que deixaram de fazer alguma tarefa é bem inferior em relação aos outros grupos. Além disso, a média de acessos ao ambiente e aos recursos disponíveis é bem maior. Como citado anteriormente, essa é uma disciplina inicial, isso pode ser um fator para um número tão expressivo de estudantes no grupo regular.

Na Unidade 2 - Base 2, houve um desequilíbrio no quantitativo de alunos entre os grupos (TABELA 2). O grupo bom tem 2 vezes mais estudantes que o ótimo. O número de participantes do grupo bom que fizeram as tarefas propostas pelo tutor é maior em relação aos outros grupos (TABELA 3). No entanto, os discentes do grupo ótimo possuem maior porcentagem de participação e atividades realizadas. Já o agrupamento regular se mantém como o que precisa de mais atenção, apesar do número de integrantes ter diminuído. Um ponto importante, é que o número de estudantes bons aumentou consideravelmente, o que leva a perceber que os alunos que tinham um percentual de participação alta, ficaram mais omissos nesta aula. Além disso, apenas 34 alunos se mantiveram no grupo dos melhores alunos nas unidades 1 e 2.

No agrupamento da Unidade 3 - Base 2, nenhum integrante do grupo ótimo deixou de fazer alguma atividade e apenas cerca de 5% dos alunos deste grupo responderam o fórum próximo do prazo final, sendo a menor taxa em relação aos demais grupos. Já o grupo regular mantém o padrão apresentado até o momento de baixa interação,

Tabela 3. Média dos valores dos atributos por grupo e unidade. Fonte: O Autor

Atributos	Unidade 1			Unidade 2			Unidade 3		
	O ¹	B ²	R ³	O	B	R	O	B	R
n_quizLV	1,41	1,32	0,00	1,17	1,58	0,00	1,74	1,75	0,00
p_quizLV	1,22	1,25	3,00	1,26	2,02	3,00	1,01	1,04	3,00
a_quizLV	4,65	2,55	0,55	11,50	5,61	1,25	7,45	5,00	1,17
n_forumLV	3,20	1,47	0,74	3,15	2,14	0,89	2,94	1,39	0,83
p_forumLV	1,98	2,73	2,69	1,91	2,46	2,62	1,16	2,25	2,69
a_forumLV	14,57	7,53	3,50	20,97	10,00	4,80	16,34	8,21	5,16
p_tarefaLV	1,44	1,64	2,95	1,65	1,73	2,44	1,11	1,43	1,90
n_tarefaLV	1,12	0,88	0,04	1,65	1,37	0,59	2,83	2,35	1,21
a_tarefaLV	13,25	6,46	0,67	22,64	10,63	2,92	13,15	8,44	1,60
a_wikiLV	20,47	4,32	0,88	31,32	11,85	1,70	22,15	6,47	1,83
n_wikiLV	5,38	0,49	0,06	7,15	2,58	0,26	4,83	0,80	0,34
p_wikiLV	1,82	2,80	2,95	2,00	2,80	2,84	1,77	2,68	2,79
a_paginas	1,60	0,85	0,27	3,23	1,43	0,60	1,98	1,22	0,50
a_url	1,61	1,04	0,37	2,76	1,64	0,74	1,97	1,38	0,74
a_curso	80,31	30,89	8,23	192,67	75,12	20,13	179,20	77,44	19,34
a_pastas	2,99	1,54	0,67	4,21	2,42	1,09	17,97	9,81	2,79
c_forumLV	2,15	1,13	0,65	2,38	1,80	0,61	2,65	1,23	0,62
faltas	19,63	32,30	88,10	16,15	20,46	76,33	7,35	19,44	63,22

¹Ótimo ²Bom ³Regular

participação e entrega de atividades. Nenhum participante fez qualquer quiz proposto durante a disciplina e o prazo de entrega de atividade é sempre próximo de 3, ou seja, próximo do prazo final. Além de 89% dos alunos não terem realizado o wiki e o nível de acessos aos recursos ser o mais baixo em relação aos demais.

Tabela 4. Tabela de classificação em relação a aprovação. Fonte: O Autor

Base 1									
	Unidade 1			Unidade 2			Unidade 3		
	Apr ¹	Rec ²	Rep ³	Apr	Rec	Rep	Apr	Rec	Rep
Ótimo	100%	0%	0%	94%	3%	3%	80%	0%	20%
Bom	92%	4%	4%	87%	9%	4%	95%	3%	2%
Regular	58%	11%	31%	25%	9%	66%	15%	19%	66%
Base 2									
	Unidade 1			Unidade 2			Unidade 3		
	Apr	Rec	Rep	Apr	Rec	Rep	Apr	Rec	Rep
Ótimo	99%	0%	1%	89%	7%	4%	95%	1%	4%
Bom	67%	18%	15%	88%	5%	7%	52%	23%	25%
Regular	30%	18%	52%	24%	20%	56%	3%	0%	97%

¹Aprovado ²Recuperação ³Reprovado

Na Tabela 4 é apresentado uma correlação entre os grupos de cada unidade com a classificação de aprovação dos estudantes. Observa-se que nas Unidades 1 e 2, a taxa de

acerto do grupo ótimo foi praticamente acima dos 90% em relação aos aprovados. Já os alunos classificados como reprovados ou de recuperação se dividem entre os grupos bom e regular, somando um percentual de, aproximadamente, 70% no último grupo. As taxas de acerto na Unidade 3 são altas, pois a mesma engloba todas as atividades.

5. Conclusões

No presente artigo, os dados de interação e visualização dos alunos no ambiente Moodle foram analisados pelo método de agrupamento hierárquico e três grupos foram identificados, consistindo de estudantes que exibiram padrões de comportamento semelhantes. No experimento com a Base 1, os alunos que mais acessam o ambiente e seus recursos possuem melhores notas e mais chances de aprovação, no entanto, isso não é pré-requisito pois cerca de 15% dos estudantes considerados regular foram aprovados no fim da disciplina, isso mostra que apesar de terem poucos acessos, estes são de qualidade.

Já no experimento com a Base 2, observa-se os alunos que entregam atividade antes do prazo possuem mais chances de serem aprovados do que os que entregam já no final do prazo e os que fazem todas as atividades, possuem média final maior que os demais. A atividade que aumentou a média com o passar das unidades foi o quiz e a tarefa (TABELA 3). Dessa forma, pode-se inferir que elas tiveram um maior grau de importância em relação aos demais. Observa-se, também, que na Unidade 2, os alunos mantiveram um padrão mais disperso com pouca participação no ambiente. Caso essa unidade ainda estivesse ativa seria possível fazer alguma intervenção pedagógica, tentando evitar esse cenário na próxima aula. Essas intervenções diminuem a distância transacional que existem no AVA.

Analisando ambos experimentos, observa-se que padrão de comportamento entre as unidades se manteve, ou seja, os alunos classificados nos grupos apresentavam o mesmo perfil de interação. Além disso, o grupo ótimo une os alunos que não precisam de uma atenção especial com uma taxa de erro pequena, isso faz com que o tutor possa focar nos estudantes do grupo bom e regular.

Por fim, conclui-se que a metodologia utilizada identifica os alunos que estão mais dispersos e que precisam de intervenções. Assim, o tutor já no início da disciplina pode aplicar o agrupamento e fazer um acompanhamento mais próximo dos alunos no grupo bom e regular, tentando identificar quais as dificuldades enfrentadas, diminuindo os índices de reprovação e evasão. Como trabalhos futuros, pretende-se utilizar outros métodos de clusterização e aplicar a metodologia em uma disciplina em andamento.

Referências

- ABED (2017). *Censo EAD.BR (2016)*. InterSaber, Curitiba, 1 edition.
- Akçapınar, G., Altun, A., and Cosgun, E. (2014). Investigating students' interaction profile in an online learning environment with clustering. In *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*, pages 109–111. IEEE.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Brazilian Journal of Computers in Education*, 19(02):03.

- dos Santos, D. C. V.-B. and Falcão, T. P. (2017). Acompanhamento de alunos em ambientes virtuais de aprendizagem baseado em sistemas tutores inteligentes. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1267.
- Gotardo, R., Cereda, P. R. M., and Junior, E. R. H. (2013). Predição do desempenho do aluno usando sistemas de recomendação e acoplamento de classificadores. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 24, page 657.
- Gottardo, E., Kaestner, C., and Noronha, R. V. (2012). Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 23.
- Kampff, A. J. C., Reategui, E. B., and Lima, J. V. d. (2008). Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. *RENOTE: revista novas tecnologias na educação [recurso eletrônico]. Porto Alegre, RS.*
- Leite, E. A. M., Sales, G. L., Joye, C. R., and de Oliveira, C. A. M. (2015). Formação de tutores na dead/ifce: Competências e habilidades necessárias. *Conexões-Ciência e Tecnologia*, 10(2):53–60.
- Losada, M. and Heaphy, E. (2004). The role of positivity and connectivity in the performance of business teams: A nonlinear dynamics model. *American Behavioral Scientist*, 47(6):740–765.
- McLaughlin, M. (2013). *MySQL Workbench: Data Modeling & Development*. McGraw Hill Professional.
- Młynarska, E., Greene, D., and Cunningham, P. (2016). Indicators of good student performance in moodle activity data. *arXiv preprint arXiv:1601.02975*.
- Moore, M. G. (2013). The theory of transactional distance. In *Handbook of distance education*, pages 84–103. Routledge.
- Passerino, L. and Gerling, C. A. (2013). Gerenciamento em ambientes virtuais de educação a distância. *RENOTE*, 3(2).
- Sales, G. (2010). *Learning Vectors (LV): um modelo de avaliação da aprendizagem em EaD online aplicando métricas não-lineares*. PhD thesis, Tese Doutorado. Departamento de Engenharia de Teleinformática. Universidade Federal do Ceará. 2010. 239f.
- Seixas, C. A., Mendes, I. A. C., Costa, S. d. G., Mazzo, A., Trevizan, M. A., Martins, J. C. A., et al. (2012). Ambiente virtual de aprendizagem: estruturação de roteiro para curso online. *Revista Brasileira de Enfermagem*, 65(4):660–666.
- Talavera, L. and Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In *Workshop on artificial intelligence in CSDL. 16th European conference on artificial intelligence*, pages 17–23.
- Wilks, D. S. (2011). Cluster analysis. In *International geophysics*, volume 100, pages 603–616. Elsevier.