

Um Ambiente Inteligente de Avaliação de Comportamentos de Tutores e Turmas no Ambiente Virtual de Aprendizagem Moodle

Rafael Castro de Souza¹, Francisco Milton Mendes Neto¹, Araken de Medeiros Santos¹, Laysa Mabel de Oliveira Fontes², Efraim N. H. D. Rodrigues³, Ricardo Alexandro de Medeiros Valentim²

¹ Programa de Pós-Graduação em Ciência da Computação – UERN/UFERSA
Av. Francisco Mota, 572 - Bairro Costa e Silva – Mossoró, RN - Brasil

² Programa de Pós-Graduação em Engenharia Elétrica e de Computação – UFRN
Campus Universitário - BR 101 - Lagoa Nova - Natal, RN - Brasil

³ Núcleo Tecnológico em Engenharia de Software (NTES) – UFERSA
Av. Francisco Mota, 572 - Bairro Costa e Silva – Mossoró, RN - Brasil

{rafaelcastro,miltonmendes,araken,mabel.fontes}@ufersa.edu.br,
ricardo.valentim@ufrnet.br, efrainnaassom@gmail.com

Abstract. *The advantages granted by the use of technology has been one of the reasons for its broad usage in multiple sectors, like, for example, in the medical, industrial, military, among others. In education, the use of Internet and computers, has facilitated access to educational content, besides, it has been improved distance education courses management through Web applications. While these applications support the creation of forums, messaging, files attachment, and a range of other tasks, they are not always capable to evaluate mentors and students interactions. Within this perspective, this paper presents a device to evaluate distance tutors activities, as well as their respective groups, in order to enable, real time monitoring of this interactions, by pedagogical sectors, so, risk situations, such as course abandonment, reproofs and low participation, can be reduced.*

Resumo. *As facilidades promovidas pelo uso da tecnologia têm sido um dos motivos para seu grande uso em vários setores, como, por exemplo, na área médica, industrial, militar, entre outros. Na educação, o uso do computador em conjunto com a Internet facilitou o acesso a conteúdos educacionais, além do gerenciamento de cursos a distância por meio de aplicações para Web. Muito embora essas aplicações suportem a criação de fóruns, envio de mensagens, inserção de arquivos, além de outras atividades, nem sempre estas dispõem de meios para avaliação das interações entre os estudantes e seus mentores. Diante dessa perspectiva, o presente trabalho apresenta uma ferramenta capaz de avaliar as atividades dos tutores a distância, bem como de suas respectivas turmas, a fim de permitir que os setores pedagógicos possam acompanhar estas turmas, em tempo real, de modo que situações de risco, tais como, abandono de curso, reprovações e baixa participação, possam ser atenuadas.*

1. Introdução

O Ensino a Distância (EaD) surgiu no ano de 1728, quando o novo método de ensino de Caleb Philips, através de aulas por correspondência, foi anunciado. Ao longo do tempo, o avanço da tecnologia permitiu que outros meios fossem também utilizados no EaD, como o rádio e até mesmo a TV Educacional, que teve seu início nas décadas de 60 e 70 no Brasil (Litto e Formiga, 2009). No entanto, esses meios de comunicação não permitem um dinamismo entre os alunos e os tutores, o que pode gerar uma desmotivação por parte do aluno.

O EaD, na década de 90, foi marcado pela integração do computador e das estações de multimídias no processo de ensino e aprendizagem (Faria e Salvadori, 2010), o que acabou gerando novas expectativas, pois possibilitou o acesso a informações sistematizadas e uma interação entre os alunos e os tutores. Para Litto e Dias (2009), houve uma maximização nas vantagens do EaD, pois a utilização dos novos meios de comunicação, técnicas e metodologias de ensino que obedecem padrões específicos geram uma maior qualidade no processo de aprendizagem.

Os sistemas de *software* concebidos para apoiar o ensino por meio da administração de um ou mais cursos, com um ou mais alunos, são denominados *Learning Management Systems* (LMS). Por meio deles, torna-se possível realizar uma gama de atividades que antes, na maior parte das vezes, somente era possível de serem realizadas de forma presencial, como, por exemplo, a aplicação de atividades, a divulgação de notas, a entrega de trabalhos, a comunicação entre alunos e professores, dentre outras (Kats, 2013).

Devido sua praticidade, os LMS vêm sendo utilizados cada vez mais por instituições de ensino. E essa utilização torna-se ainda mais acentuada nas modalidades de EaD, visto que o gerenciamento de cursos, matrículas, lançamento de notas e a própria comunicação entre professores e alunos podem ser feitos em qualquer lugar, desde que se tenha um aparato computacional conectado à Internet, como, por exemplo, um computador, *tablet*, *smartphone* ou qualquer outro dispositivo congênere.

Segundo o último censo realizado pela Associação Brasileira de Educação a Distância (ABED), o número de matrículas em 2014 (desde o ensino fundamental até a pós-graduação: *stricto sensu* - doutorado) somaram 3.868.706 matrículas, com 519.839 (13%) nos cursos regulamentados totalmente a distância, 476.484 (12%) nos cursos regulamentados semipresenciais ou disciplinas de cursos presenciais e 2.872.383 (75%) nos cursos livres (Associação Brasileira de Educação a Distância, 2014).

Por outro lado, o maior obstáculo enfrentado em 2014 nesta modalidade de ensino foi a evasão dos estudantes, seguido pela resistência dos educadores à modalidade de EaD, combinada aos desafios organizacionais de uma instituição presencial que passa a oferecer essa modalidade de ensino. De modo que, em todos os tipos de cursos analisados, a evasão dos alunos se concentra na faixa de até 25%, sendo que uma das suas causas é a falta de participação nos cursos (Associação Brasileira de Educação a Distância, 2014).

Uma vez que os LMS tendem a registrar notas de alunos, horários de acessos, visualizações e participações em fóruns de discussão, quantidade de exercícios resolvidos e demais ações, tanto de alunos quanto de professores, estes contêm amplas informações acerca do processo de ensino. Portanto, seus bancos de dados tornam-se ideais para a aplicação da *Learning Analytics* (LA), cujo objetivo é a interpretação de

uma ampla gama de dados, produzidos pelos estudantes, professores e tutores, a fim de avaliar o progresso acadêmico, prever o desempenho futuro e identificar possíveis problemas de aprendizagem (Cooper, 2012).

Vale salientar que uma ferramenta de LA não deve limitar-se somente a recuperação da informação e sua exibição, além disso, é necessário analisar esses dados de modo que conhecimento possa ser extraído a partir da análise a fim de permitir a tomada de decisões pedagógicas com segurança (da Silva e Imran, 2015), dessa forma, o uso de técnicas de análise de dados, tais como as técnicas estatísticas, ou do campo da Inteligência Artificial, são intrínsecas ao processo da LA.

Diante dessa perspectiva, este trabalho apresenta uma ferramenta capaz de classificar turmas de acordo com a taxa de participação e com as ações dos alunos e tutores, objetivando promover a participação efetiva, além de atenuar as chances de evasão dos alunos e oferecer um suporte aos tutores na adaptação dessa modalidade de ensino.

Este artigo está organizado da seguinte forma: a Seção 2 descreve a fundamentação teórica da *Learning Analytics*, a Seção 3 aborda conceitos da área de Aprendizado de Máquina, a Seção 4 expõe alguns trabalhos relacionados com a temática deste artigo, na Seção 5, é descrita a arquitetura da referida ferramenta, bem como os resultados obtidos, e, por fim, a Seção 6 expõe as considerações finais.

2. Learning Analytics

A primeira conferência sobre *Learning Analytics* (LA) ocorreu em 2011, na cidade de Banff, Alberta, Canadá (Siemens, 2011). Nessa conferência, a LA foi definida como sendo a medição, coleta, análise e comunicação de dados sobre os alunos e os seus contextos, para fins de compreensão e otimização da aprendizagem nos ambientes em que esse processo ocorre (Siemens et. al., 2011).

Já Lias e Elias (2011) definem a LA de uma forma mais sucinta, como sendo um campo emergente em que ferramentas de análise são utilizadas a fim de melhorar a aprendizagem e a educação. Outros autores, como Johnson *et al.* (2011), por exemplo, definem a LA da seguinte forma: “*Learning Analytics* refere-se à interpretação de uma ampla gama de dados produzidos pelos estudantes e coletados pelo seu interesse, a fim de avaliar o progresso acadêmico, prever o desempenho futuro e identificar possíveis problemas”.

Ainda no mesmo ano da primeira conferência sobre a LA, foi criada a SOLAR (*Society and Learning Analytics Research*), que é uma rede interdisciplinar de pesquisadores internacionais que vêm pesquisando sobre o papel e o impacto que essas análises podem gerar sobre o processo de ensino e aprendizagem (Siemens et. al., 2011). A SOLAR faz parte da organização da Conferência Internacional sobre *Learning Analytics & Knowledge* (LAK), atuando também no suporte ao lançamento de várias iniciativas no apoio à pesquisa colaborativa sobre a área (Siemens et. al., 2011).

Outra área de pesquisa bastante relacionada com a LA é a Mineração de Dados Educacionais (do inglês, *Educational Data Mining* - EDM). Porém, muito embora essas duas áreas trabalhem com grandes volumes de dados educacionais, elas são consideradas duas áreas de pesquisa distintas (de Faria, 2014). A seguir são apresentadas as diferenças entre essas duas áreas, segundo Siemens e Baker (2012) e Romero e Ventura (2013):

- Objetivo: De forma sucinta, a EDM procura novos padrões nos dados, a partir do desenvolvimento de novos algoritmos e/ou novos modelos. Já a LA aplica modelos e algoritmos já existentes nos sistemas de ensino e aprendizagem;
- Ênfase: A LA tem uma ênfase maior no detalhamento, apresentação e descrição dos resultados e dados, enquanto que a EDM foca mais nas técnicas utilizadas na mineração em si;
- Tipo de Descoberta: Na LA, utiliza-se da descoberta automática como forma de ajudar na tomada de decisões. Já na EDM, valoriza-se mais a criação ou melhoria do método de descoberta automatizada da informação.

Sendo assim, diante do exposto, a ferramenta proposta pelo presente trabalho situa-se no campo investigativo da LA, tendo em vista que a mesma utiliza métodos de análise de dados com o propósito de ajudar na identificação de problemas de baixa participação das turmas e também na análise do desempenho dos tutores, objetivando melhorar a tomada de decisão destes e de todos os demais responsáveis pelo processo de ensino e aprendizagem.

No entanto, uma ferramenta de LA não deve apenas estar limitada à recuperação de dados de um LMS e à sua apresentação para um professor, por exemplo, mas também deve contribuir na organização, análise e interpretação das informações em si, para que o professor possa tomar medidas pedagógicas com segurança. A seção a seguir apresenta os conceitos da área de Aprendizado de Máquina, bem como expõe os conceitos da técnica que foi empregada neste trabalho.

3. Aprendizado de Máquina

A área de Aprendizado de Máquina (AM) é uma área da Ciência da Computação, comumente considerada uma subárea da Inteligência Artificial, cujo objetivo de pesquisa é o desenvolvimento de métodos e técnicas a fim de encontrar padrões, regularidades ou conceitos em conjuntos de dados (Goldschmidt e Bezerra, 2015).

Conforme Faceli *et al.* (2011), existem vários problemas reais cujas técnicas de AM podem ser aplicadas, dentre as quais podem ser citadas: (i) predição da taxa de cura de pacientes; (ii) reconhecimento de palavras; (iii) identificação de fraudes em cartões de crédito; e (iv) diagnóstico de câncer ou outros tipos de doenças.

Ainda conforme Faceli *et al.* (2011), as tarefas de aprendizado de máquina se dividem em duas categorias, que são a predição e a descrição. A predição consiste em encontrar uma fórmula ou gerar um modelo que seja capaz de prever ou rotular um dado a partir de um conjunto de dados para treinamento. Os algoritmos que pertencem a essa categoria são denominados de algoritmos de aprendizado supervisionado. O termo supervisionado advém da presença de um rótulo ou valor de saída existente nos dados que foram utilizados para treinamento do conjunto. Dessa forma, as técnicas de aprendizado supervisionado usam o valor do rótulo de saída do conjunto de dados de treinamento para avaliar a capacidade da hipótese induzida de predizer os valores de saída para novos exemplos.

Já nas tarefas de descrição, o objetivo é explorar o conjunto de dados. Os algoritmos pertencentes a essa categoria são denominados de algoritmos de aprendizado não supervisionado. O termo não supervisionado vem da inexistência de um rótulo de

saída nos dados do conjunto de treinamento. Dessa forma, uma vez que não há rótulo de saída, esses algoritmos buscam explorar os dados a partir de sua regularidade.

Tendo em vista que não há rótulo de saída nos dados comportamentais dos alunos e tutores, as técnicas que se adequam ao enfoque deste trabalho, que é a classificação de comportamentos e ações das turmas e dos tutores do ensino a distância, são as técnicas descritivas.

Existem várias técnicas de aprendizado não supervisionado, como, por exemplo, agrupamento, sumarização, associação, cada uma com suas particularidades, sendo que, para cada técnica, existem inúmeros algoritmos diferentes. Dessa forma, para a escolha da técnica que seria empregada na ferramenta proposta, foram levados em consideração três critérios: (i) desempenho (verificação da taxa de erro); (ii) tempo de execução; e (iii) simplicidade para interpretação dos resultados.

Após um estudo comparativo, a técnica que apresentou o melhor resultado, segundo os critérios supracitados, foi o K-Means. A subseção a seguir apresenta a fundamentação teórica dessa técnica.

3.1 K-Means

O algoritmo K-Means, proposto por J. MacQueen (MacQueen, 1967), foi concebido para particionar uma população n -dimensional em k conjuntos, com base em uma amostra. Primeiro, o algoritmo inicia com a escolha de k elementos que formarão os *clusters* iniciais. O número k refere-se ao número de *clusters* definidos. Por exemplo, para $k = 2$, o algoritmo agrupará os elementos em dois *clusters*. A escolha inicial da posição dos *clusters* pode obedecer aos seguintes critérios: (i) por meio da seleção das k primeiras instâncias; (ii) selecionando k instâncias de forma aleatória; e (iii) por meio da seleção de k instâncias que possuam alto grau de dissimilaridade.

Após terem sido escolhidos os *clusters* iniciais, são calculadas as distâncias dos demais elementos em relação aos *clusters*, de modo que os elementos que possuírem a menor distância serão agrupados ao *cluster*. Em seguida, é recalculado o centróide deste *cluster*, levando em conta os novos elementos que foram agrupados. Esse processo é repetido até que todas as instâncias pertençam a um *cluster* (Wu, 2012).

A seguir serão apresentados alguns trabalhos relacionados a esta proposta.

4. Trabalhos Relacionados

Arnold e Pistilli (2012) propuseram um sistema, denominado *Course Signals* (CS), que fornece um *feedback* significativo para o aluno, com base em modelos preditivos. A ideia desse sistema é utilizar a riqueza de dados encontrada em uma instituição de ensino, e, por meio de análises sobre os dados extraídos, determinar, em tempo real, quais os momentos em que esses alunos podem estar em risco de reprovação.

Dimopoulos *et al.* (2013) desenvolveram um *plugin* para o Moodle (versão 2.2+), denominado LAe-R tool, que é capaz de analisar uma série de competências dos alunos, como, por exemplo, participação em fóruns ou *chats*, interação com outros usuários, além de outros critérios que podem ser definidos pelo professor, no intuito de ajudá-lo na identificação de elementos que possam estar associados à aprendizagem dos alunos. Uma vez definidos esses critérios, o *plugin* retorna os dados dos alunos correspondentes às características definidas pelo professor, e, em seguida, forma uma

tabela de apresentação para este, onde, nessa tabela, as linhas correspondem aos critérios e as colunas correspondem aos respectivos níveis de desempenho dos alunos.

Similarmente ao trabalho supracitado, Zielinski e Schmitt (2015) também apresentaram um *plugin* para o Moodle, porém, neste caso, o *plugin* permite o acesso a visualizações do conteúdo acessado, das submissões e da participação da turma por meio de gráficos.

Os fatores que diferem o presente trabalho dos demais trabalhos apresentados nesta seção são: (i) quanto ao modelo de análise: uso do modelo descritivo (aprendizado não supervisionado); (ii) quanto aos atributos analisados: foco em ações comportamentais; (iii) alvos da análise: alunos e tutores a distância; e (iv) quanto à forma de avaliação: avalia os atributos a partir da técnica K-Means. Além disso, apresenta um gráfico informativo exibindo a posição relativa dos atributos analisados em comparação com o melhor desempenho já obtido até o momento para os outros atributos de outras turmas.

Na próxima seção, será descrita a arquitetura da ferramenta proposta, bem como os resultados obtidos.

5. Arquitetura e Resultados da Ferramenta de Avaliação

A ferramenta foi construída utilizando a linguagem de programação JAVA (Oracle, 2016). Para o uso do algoritmo K-Means, foi utilizada a biblioteca WEKA (*Waikato Environment for Knowledge Analysis*) (WEKA, 2016), que é uma ferramenta feita em JAVA que contém uma coleção de algoritmos de aprendizado supervisionado e não supervisionado (WEKA, 2016). E para a construção do Core III, foi utilizada a linguagem de programação *Javasever Faces* (Oracle, 2016), em conjunto com a API *PrimeFaces* (PrimeFaces, 2016).

A ferramenta proposta contém três núcleos (*cores*) de processamento de dados que são descritos nas subseções a seguir.

5.1 Núcleo de Extração de Dados (Core I)

O primeiro núcleo é responsável pela extração das informações no banco de dados. Vale salientar que, neste trabalho, os dados foram cedidos pela Secretaria de Educação a Distância (SEDIS), pertencente à Universidade Federal do Rio Grande do Norte (UFRN). Essa instituição usa o Moodle como LMS. Portanto, todas as informações que compõem o *dataset* foram extraídas desse LMS.

Ao todo, foram extraídos dados de 62 turmas, pertencentes a 10 cursos de graduação, que são: (i) Bacharelado em Administração Pública; (ii) Licenciatura em Ciências Biológicas; (iii) Licenciatura em Educação Física; (iv) Licenciatura em Física; (v) Licenciatura em Geografia; (vi) Licenciatura em Letras; (vii) Licenciatura em Matemática; (viii) Licenciatura em Pedagogia; (ix) Licenciatura em Química; e (x) Licenciatura em História. A base de dados histórica cedida corresponde aos anos de 2012 a 2013. No total, o *dataset* foi composto por informações de 2.227 alunos e 38 tutores a distância.

Este núcleo extrai as seguintes informações sobre os tutores e turmas:

- Tutores: (i) número de tópicos criados; (ii) média de postagens em tópicos; (iii) taxa de visualização em fóruns; (iv) taxa de visualização em tópicos; (v) número

de *chats* criados; (vi) média de postagens em *chats*; (vii) número de cliques; (viii) número de páginas criadas; (ix) número de URLs criadas; e (x) número de arquivos criados.

- Turma: (i) taxa de participação nos questionários; (ii) tempo médio para finalização dos questionários; (iii) média de tópicos criados (iv) média de postagens em tópicos por alunos; (v) média de visualização em fóruns por alunos; (vi) média de visualização em tópicos por alunos; (vii) taxa de submissão de atividades da turma; (viii) média de postagens em *chats* por aluno; (ix) média de cliques dos alunos da turma; (x) média de páginas visualizadas pelos alunos; e (xi) média de arquivos visualizados pelos alunos.

Após a extração dos dados, esse núcleo armazena os dados no formato .arff, utilizando a mesma estrutura de armazenamento que é utilizada pelo *software* WEKA. O *dataset* gerado pelo Core I servirá como entrada de dados para o núcleo de processamento de dados que será descrito a seguir.

5.2 Núcleo de Processamento de Dados (Core II)

O núcleo de processamento de dados tem o objetivo de processar as informações que foram extraídas pelo Core I. Ao todo são realizados dois tipos de processamento no Core II: (i) criação e treinamento do algoritmo K-Means; e (ii) cálculo da posição relativa dos atributos analisados.

A criação e treinamento do algoritmo de classificação consistem em criar, para cada atributo analisado, a sua própria rede de classificação. Em outras palavras, cada atributo do tutor e da turma, que será analisado, terá sua própria rede de classificação, que será classificada por meio da técnica K-Means, utilizando-se do critério da distância euclidiana entre os centróides, para $k = 3$, ou seja, serão criados três agrupamentos. A partir disso, será possível identificar se o atributo em questão pertence a categoria dos que tiveram baixo desempenho, desempenho moderado ou um bom desempenho. Vale ressaltar que essa rede de classificação será treinada por meio das instâncias do *dataset* resultante do processamento do Core I.

Além da classificação propriamente dita, que tem por objetivo classificar cada um dos atributos por meio da técnica de agrupamento, o Core II calcula também a posição relativa de cada atributo avaliado em comparação com os dados das demais turmas. Para isso, é utilizada a técnica de normalização min-max, que é definida pela Equação 1 (Faceli et al., 2011).

$$v' = \frac{(v - \min)}{(\max - \min)} \times (n\max - n\min) + n\min \quad (1)$$

Onde:

- i) v é o valor a ser normalizado; ii) \min é o valor mínimo encontrado no *dataset*; iii) \max é o valor máximo encontrado no *dataset*; iv) $n\max$ é o valor máximo do novo intervalo; v) $n\min$ é o valor mínimo do novo intervalo.

Desse modo, além da classificação, que irá permitir uma avaliação de cada atributo do tutor e da turma, por meio da técnica de agrupamento de dados, será ainda

exposto um gráfico informando a posição relativa do desempenho da turma em comparação com as demais turmas.

5.3 Núcleo de Apresentação de Dados (Core III)

O núcleo de apresentação de dados (Core III) é responsável pela exibição das informações processadas no Core II, e também por permitir a interação do usuário com a ferramenta. A Figura 1 apresenta a listagem dos cursos a serem avaliados.

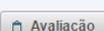
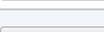
ID	Curso	Analisar Turma
5	Licenciatura em Geografia	
8	Licenciatura em Química	
10	Bacharelado em Administração Pública	
85	DISCIPLINA ESTÁGIO SUPERVISIONADO II (ENSINO DE FÍSICA E CIÊNCIAS) - CFD0022_2013.1	
86	DISCIPLINA ESTÁGIO SUPERVISIONADO III (ENSINO DE FÍSICA E CIÊNCIAS) - CFD0023_2013.1	
92	DISCIPLINA ESTÁGIO SUPERVISIONADO EM ENSINO DE QUÍMICA II - CQD0002_2013.1	

Figura 1. Listagem dos cursos para avaliação.

Após a escolha do curso a ser avaliado, os seus dados são extraídos em tempo real (Core I), e, em seguida, são processados no Core II, e, por fim, as respectivas informações de classificação e desempenho relativo são apresentadas (Core III), conforme ilustra a Figura 2.

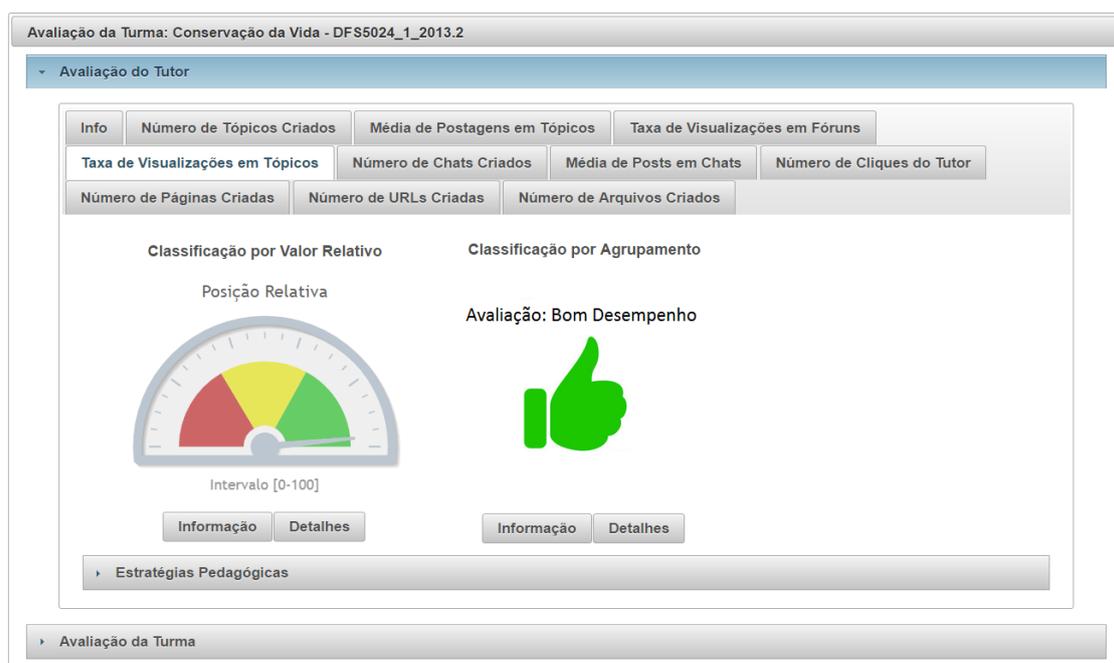


Figura 2. Informação indicativa da avaliação do atributo Taxa de Visualizações em Tópicos pelo tutor mediante o uso da técnica *min-max* e aprendizado não supervisionado.

6. Considerações Finais

O presente trabalho apresentou uma ferramenta capaz de avaliar atributos de tutores e turmas na modalidade de ensino a distância, mais especificamente das turmas ofertadas no LMS Moodle. Para a construção do *dataset*, cujo objetivo é servir como entrada para o treinamento da rede de classificação, foram obtidos dados de 62 turmas do ensino a distância, onde o *dataset* reuniu informações comportamentais de 2.227 alunos e 38 tutores a distância. Ao todo, a ferramenta é capaz de avaliar dez comportamentos dos tutores e onze comportamentos das turmas, totalizando uma avaliação de 21 atributos comportamentais.

Para a avaliação comportamental foi utilizado a técnica K-Means, que pertence à categoria do aprendizado de máquina não supervisionado e realiza uma classificação por meio de agrupamento de dados. A escolha dessa técnica se deu pelo bom desempenho apresentado quanto à taxa de acertos, o tempo de execução e sua simplicidade para interpretação dos resultados.

Com a ferramenta proposta se torna possível que os coordenadores de cursos, os setores pedagógicos, os professores e os tutores possam refletir sobre os desempenhos obtidos. Além disso, a ferramenta permite o monitoramento de turmas, em tempo real, de modo que estratégias pedagógicas possam ser traçadas para melhorar o ensino, promover a participação efetiva e minimizar as evasões, bem como as reprovações nas disciplinas.

7. Referências Bibliográficas

- Associação Brasileira de Educação a Distância (ABED). Disponível em <http://www.abed.org.br/censoead2014/CensoEAD2014_portugues.pdf> Acesso em: 15 de março de 2016.
- Arnold, K.E., Pistilli, M.D. Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the 2nd International Conference on Learning Analytics & Knowledge, 2012.
- Cooper, A. What is analytics? Definition and essential characteristics. CETIS Analytics Series, 1(5), 1-10, 2012.
- da Silva, J. M. C., Imran, H. Um estudo sobre as variáveis para predição de alunos não concluintes em cursos suportados por Ambientes Virtuais de Ensino e Aprendizagem. *RENOTE*, 13(2), 2015.
- de Faria, S. M. S. M. L. Educational Data Mining e Learning Analytics na melhoria do ensino online. Dissertação para obtenção do Grau de Mestre em Estatística, Matemática e Computação, Universidade Aberta, 2014.
- Dimonoulos, I., Petronoulou, O., Boloudakis, M., & Retalis, S. Using Learning Analytics in Moodle for assessing students' performance. In *2 nd Moodle Research Conference*, 2013.
- Faceli K., Lorena A. C., Gama J., de Carvalho A. C. P. L. F. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC, 2011.
- Faria A. A., Salvadori A. Educação a Distância e seu movimento histórico no Brasil. *Revistas das Faculdades Santa Cruz, Curitiba*. v. 8, n. 1. 2010.
- Goldschmidt R., Bezerra E. Data mining : conceitos, técnicas, algoritmos, orientações e aplicações. Elsevier Brasil, 2015.
- Johnson, L., Smith, R., Willis, H., Levine, A., Haywood, K. The 2011 Horizon Report. The New Media Consortium, 2011.
- Kats, Y. Learning Management Systems and Instructional Design: Best Practices in Online Education. Idea Group Inc (IGI), 2013.

- Lias, T. E., & Elias, T. Learning Analytics: The Definitions, the Processes, and the Potential, 2011.
- Litto F. M., Formiga M. Educação a Distância: o estado da arte. São Paulo: Person Education do Brasil. 2009.
- MacQueen, J. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281-297, 1967.
- Oracle. Disponível em <<http://www.oracle.com/technetwork/java/javaee/javaserverfaces-139869.html>> Acesso em 13 de março de 2016.
- PrimeFaces. Disponível em < <http://www.primefaces.org/>> Acesso em 13 de março de 2016.
- Romero, C., Ventura, S. Data Mining in Education. Data Mining Know Discov, 3, 12-27, 2013.
- Siemens G. LAK'11 1st Internacional Conference on Learning Analytics and Knowledge. Disponível em < <https://tekri.athabasca.ca/analytics/>> Acesso em: 10 de junho de 2015.
- Siemens, G., Baker, RSJd. Learning analytics and educational data mining: towards communication and collaboration. Em: Proceeding of the 2nd Internacional Conference on Learning Analytics and Knowlwdge. Vancouver, British Columbia, Canada. 1-3, 2012.
- Siemens G., Gasevic D., Haythornthwaite C., Dawson S., Shum S.B., Ferguson R., Duval E., Verbert K., Baker R.S.J.d.. Open Learning Analytics: an integrated & modularized platform. Society for Learning Analytics Research. Disponível em< <http://solaresearch.org/OpenLearningAnalytics.pdf>> Acesso em: 2 de junho de 2015.
- WEKA. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/index.html>> Acesso 13 de março de 2016..
- Wu J. Advances in K-means Clustering: A Data Mining Thinking. Springer Science & Business Media, 2012
- Zielinski, F. D. C., Schmitt, M. A. R. Uma ferramenta gráfica para suporte à atividade docente no Moodle. RENOTE, 13(1), 2015.