

Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância

Emanuel Marques Queiroga^{1,2}, Cristian Cechinel¹, Ricardo Matsumura Araujo¹

¹PPGC – Universidade Federal de Pelotas (UFPel)
Pelotas – RS – Brazil

²Instituto Federal Sul-rio-grandense (IFSUL) – campus Pelotas Visconde da Graça
Pelotas – RS – Brasil

{emanuel.queiroga, ricardo}@inf.ufpel.edu.br

contato@cristiancechinel.pro.br

Abstract. *Distance Learning (DL) is a modality of education mediated by technology, which students and teachers usually interact through a Virtual Learning Environment (VLE). With the growth of this type of education arise new challenges and opportunities in learning and computation, related to a better use of available resources. One of the key features present in the DL is the availability of data generated by user interactions. This data can contain records of each user interaction, including date and time of these interactions, as well as its location and type. Thus, you can use this data to discover and model the behavior of different types of users. This article describes the initial results of a work targeted to early prediction of students' dropout in a distance learning course, using data mining about their interactions in the initial 4 weeks of the course.*

Resumo. *A Educação a Distância (EAD) é uma modalidade de ensino mediada por tecnologias em que discentes e docentes utilizam um Ambiente Virtual de Aprendizagem (AVA). Com o crescimento desta modalidade de ensino surgiram novos desafios e oportunidades pedagógicos e computacionais, relacionadas a um melhor aproveitamento dos recursos disponíveis. Um dos principais recursos presente na EAD é a disponibilidade de dados gerados pelas interações dos usuários. Estes dados podem conter registro de cada interação dos usuários, incluindo data e hora destas interações, assim como sua localização e tipo. Assim, é possível utilizar estes dados para descobrir e modelar o comportamento de diferentes tipos de usuários. O presente artigo descreve os resultados iniciais de um trabalho voltado para a predição precoce da evasão de acadêmicos de um curso de educação a distância utilizando mineração dos dados relativos as suas interações nas 4 semanas iniciais do curso.*

1. Introdução

No atual contexto social tornou-se indispensável a busca por conhecimentos e qualificação das pessoas, de forma que nos últimos anos o governo brasileiro através do Ministério da Educação e entidades de fomento tem feito uma série de investimentos em programas de educação buscando a qualificação da mão de obra produtiva no país. Considerando que o Brasil é um país de grandes dimensões, diversas cidades estão afastadas

dos grandes centros universitários e acabam ficando isoladas de programas de graduação e cursos técnicos profissionalizantes. Desta forma, uma das alternativas adotadas pelo governo federal para a expansão do acesso a educação foi a utilização da modalidade à distância (Educação a Distância - EAD), que tem como um de seus objetivos levar o ensino a estas localidades, geralmente utilizando Ambientes Virtuais de Aprendizagem (AVA) [Delano and Corrêa 2013].

O AVA é o “local virtual” onde os cursos na modalidade a distância, ou semi-presenciais, normalmente acontecem. São ambientes que utilizam plataformas especialmente planejadas para abrigar cursos. Uma das plataformas mais utilizadas no país é o Moodle¹. Nela, existem áreas para apresentação de conteúdos em diversos formatos, atividades de verificação da aprendizagem e espaços para interação síncrona, por meio de chats, e assíncrona, através de fóruns de discussão. Tratam-se de recursos que permitem a interação dos estudantes entre si e com a equipe de tutores e professores. A organização do ambiente virtual permite ao aluno um acompanhamento organizado e sistematizado daquilo que é estudado a cada semana. A recuperação da informação e dos conteúdos estudados também é um dos benefícios proporcionados por cursos à distância que utilizam AVA's [Segundo and Ramos 2005].

Entretanto, um dos principais desafios da EAD é obter a diminuição do índice de evasão, que conforme o Censo EAD (CensoEAD, 2013), variou de 18,6% em 2010 para 11,74% em 2012 com pico de 20,5% em 2011 nos cursos autorizados pelo Ministério da Educação (MEC). Barroso e Falcão [Barroso and Falcão 2004] agrupam as condições desencadeantes para evasão em 3 classificações, i) econômica - impossibilidade de permanecer no curso por questões sócio-econômicas; ii) vocacional – o aluno não se identifica com o curso; iii) institucional – abandono por fracasso nas disciplinas iniciais, inadequação aos métodos de estudo, dificuldades de relacionamento com colegas ou com membros da instituição.

Segundo Manhães [Manhães et al. 2011], a detecção precoce de grupos de alunos com risco de evasão é uma condição importante para reduzir o problema da evasão, uma vez que um tratamento mais adequado pode ser ofertado a ele. Ainda segundo Manhães, atualmente o processo de identificação é manual, subjetivo, empírico e sujeito a falhas, pois depende primordialmente da experiência acadêmica e do envolvimento dos docentes. Geralmente, estes desempenham inúmeras atividades, portanto é difícil acompanhar e reconhecer as necessidades de cada aluno e identificar aqueles alunos que apresentam risco de evasão.

Diante destes fatores, a utilização da *Educational Data Mining* (EDM) se faz relevante. A EDM é uma área de pesquisa recente e que tem como principal objetivo o desenvolvimento e aplicação de técnicas de mineração de dados na exploração de conjuntos de dados coletados em ambientes educacionais. Atualmente a EDM vem se estabelecendo como uma forte e consolidada linha de pesquisa que possui grande potencial para melhorar a qualidade do ensino [Baker et al. 2011]. Essa área pode ajudar as instituições a criarem modelos de predição que tenham condições de avaliar as chances de um determinado acadêmico se evadir.

Com a análise dos dados de interações gerados pelo AVA e o auxílio da EDM,

¹<https://moodle.org/>

técnicas de mineração de dados e aprendizagem de máquina, é possível a avaliação da probabilidade de um usuário evadir. Com isso, é possível o tratamento diferenciado entre os alunos, dedicando formas de auxílio diferenciadas a um determinado aluno que esteja com uma probabilidade maior de evasão.

Considerando os aspectos apresentados, este trabalho tem como objetivo aplicar técnicas de mineração de dados e aprendizagem de máquina em dados disponíveis da EAD de uma instituição, avaliando a possibilidade e eficácia em identificar precocemente alunos em risco de evasão. É foco do trabalho utilizar apenas contagem de interações ao longo do tempo (i.e. séries temporais), uma vez que esta é uma métrica facilmente generalizável para outras plataformas e abordagens de ensino, em contraste com abordagens que são extremamente específicas (e.g. utilizando tipos de interações que não necessariamente existem em todas as plataformas ou são utilizadas em todas as execuções de cursos).

2. Trabalhos Relacionados

Diversos trabalhos procuram modelar comportamentos de alunos de EAD com a intenção de realizar previsões sobre estes, utilizando uma variedade de técnicas e bases de dados, em geral com bons resultados.

Pesquisas disponíveis sobre o tema como a de Gotardo e Hruschka [Gotardo et al. 2013], que utiliza Sistemas de Recomendação e Acoplamento de Classificadores utilizando como dados notas atribuídas a conteúdos anteriores, onde obteve resultados de 80% de acurácia com classificadores Naive Bayes na predição de desempenho de alunos.

Sandro Rigo [Rigo 2014] demonstrou técnicas com taxa de acerto na predição da evasão de 89% utilizando Educational Data Mining (EDM) e Learning Analytics(LA) aplicadas aos dados de alunos da EAD da Universidade do Vale do Rio dos Sinos (UNISINOS), assim reduzindo a evasão em 14%.

Por sua vez Wagner Cambruzzi [Cambruzzi 2014], apresenta uma taxa de acurácia média de 75,7% nos casos de evasão utilizando técnicas de EDM no pré-processamento dos dados e redes neurais na etapa de aprendizagem.

Outro trabalho relacionado que apresenta resultados interessantes é o de Douglas Detoni [Detoni et al. 2014], que busca utilizar, como o presente trabalho, unicamente contagem de interações para predição de reprovação em disciplinas de EAD, relatando resultados de até 67% de acurácia na predição utilizando redes bayesianas.

Já Manhães [Manhães et al. 2011], tem como objetivo em seu trabalho identificar precocemente alunos em risco de evasão nos cursos de graduação em Engenharia presencial da Universidade Federal do Rio de Janeiro. Para isso utiliza dados sobre o desempenho dos alunos em duas disciplinas do primeiro semestre do curso, e obteve taxas de até 80% de acurácia.

O presente trabalho se distingue dos demais ao buscar realizar predição de evasão utilizando unicamente como atributos contagens de interações ao longo do tempo. Adicionalmente, busca-se uma predição precoce da situação ao utilizar apenas dados das 4 primeiras semanas da vida acadêmica do aluno. Em contraste, a maior parte dos trabalhos utiliza um amplo espectro de atributos sem limites estritos no horizonte de predição.

3. Metodologia

Para esta pesquisa, foi realizada uma busca por artigos relacionados ao tema na base de dados "A literatura brasileira sobre mineração de dados educacionais - Relação de artigos" disponibilizada no "3º Congresso Brasileiro de Informática na Educação - CBIE 2014 - Workshop de Mineração de Dados em Ambientes Virtuais do Ensino/Aprendizagem - WMDAVDA" além de uma busca nos seguintes eventos e periódicos: Revista Brasileira de Informática na Educação (RBIE), Escola Regional de Computação Bahia - Alagoas - Sergipe (ERBASE), Simpósio Brasileiro de Informática na Educação (SBIE), Conferência Latino-Americana de Objetos e Tecnologias de Aprendizagem (LACLO) e Simpósio Brasileiro de Inteligência Artificial (SBIA).

Como palavras chaves na busca utilizou-se, EDM, MOODLE, EAD e mineração de dados educacionais. Após a separação e a leitura dos artigos relacionados, foram estudadas as particularidades de cada um dos casos, as técnicas utilizadas e os resultados obtidos. Assim buscando identificar os algoritmos mais utilizadas em casos compatíveis com a da instituição, bem como as técnicas de pré-processamento dos dados que era aplicada em cada um dos casos.

3.1. Descrição dos dados

Os dados utilizados neste trabalho, foram as interações dos alunos de EAD do de um determinado campus da instituição, que oferta quatro cursos técnicos na modalidade semi-presencial e sub-sequentes e com carga horária total de 1215 horas dívidas nas disciplinas do curso.

Estes cursos tem duração de dois anos e regime anual, sendo divididos em 2 anos letivos e ofertados nos pólos que se localizam em cidades do interior do estado. A entrada de alunos é através de vestibular, onde ingressam 50 alunos por pólos no ano.

O prazo máximo para integralização do curso é de quatro (4) anos, podendo o aluno repetir somente uma vez cada disciplina e por consequência ano, o aluno ainda tem a opção de levar até 2 disciplinas como dependência para o próximo ano e cursá-las de forma concomitante as outras disciplinas do curso. Para aprovação o aluno deverá ter média igual ou superior a seis em cada uma das disciplinas da matriz curricular. Considera-se evadido o aluno que passe um período de 365 dias sem interações com o ambiente virtual ou não efetue sua matrícula anual. Assim desligando o mesmo do curso.

Neste trabalho, optou-se por utilizar os dados de duas turmas de polos diferentes mas que tiveram o início de suas aulas na mesma data em 2013. A turma do primeiro polo é composta por 50 alunos, enquanto que a do segundo tem 34 alunos. Os dados de interações utilizados, foram das disciplinas que são inicialmente cursadas pelos alunos do curso Técnico em Administração, Fundamentos de Educação a Distância com carga horária de 45h e Informática Aplicada com 60h. Estas duas disciplinas são as primeiras interações dos alunos com a EAD e o Moodle. Valendo ressaltar ainda que a carga horária semanal dos alunos é de 15h.

Os dados foram extraídos diretamente do ambiente virtual de aprendizagem Moodle através do sistema de logs. Este sistema de logs registra todas as interações dos alunos com o ambiente. O tratamento realizado nos dados extraídos é descrito subseção 3.2.

3.2. Pré-Processamento

Na etapa de pré processamento, os dados foram separados por alunos contendo as informações referentes as interações de cada semana (semana 1, semana 2, semana 3, semana 4) e a situação final do aluno no curso (evadido ou não evadido). Nos 4 agrupamentos semanais, os dados primeiramente são separados em 7 dias a partir do dia de início das atividades. Após isto, os dados são contabilizados individualmente por número de interações do aluno naquela determinada semana. A situação final consiste na situação do aluno no final do segundo ano letivo, dizendo se até aquele momento o aluno ainda esta em situação normal ou se está evadido.

4. Ferramenta e algoritmos de mineração

Para mineração dos dados, optou-se por utilizar a ferramenta WEKA (Waikato Environment for Knowledge Analysis), na versão 3.6.11. A escolha deste software se motivou por alguns fatores como, facilidade de utilização, obtenção direta da página do desenvolvedor sem custos, número considerável de algoritmos disponíveis para utilização junto com a possibilidade de alteração dos parâmetros de execução dos mesmos e a possibilidade de fácil comparação entre os algoritmos testados.

4.1. Ambiente de Experimentação

A ferramenta de mineração disponibiliza dois ambientes para experimentação, o Weka Experiment Environment (WEE) e o Weka Explorer (WE), onde o WEE é apropriado para realizar comparações entre o desempenho de vários algoritmos de mineração de dados [Bouckaert 2010]. 2010]. O WEE permite a seleção de vários algoritmos para teste e depois analise dos resultados comparativos entre eles, assim verificando o que obteve melhores resultados estatisticamente. O outro ambiente disponível na ferramenta Weka é o Weka Explorer (WE), este ambiente permite a seleção e execução de um algoritmo classificador por vez [Bouckaert 2010].

Neste artigo optou-se pela utilização do ambiente WEE em todos os experimentos, configurando o mesmo para a opção Cross Validation com 10 folds, o que consiste em dividir a base de dados em 10 partes, sendo 9 partes utilizadas para treinamento e 1 para teste. O processamento é executado 10 vezes e o resultado final é calculado pela média de cada uma das etapas. Esta opção se torna interessante pelo fato de que os conjuntos de teste e treinamento tendem a não ser iguais por conta do número de repetição.

4.2. Algoritmos de Mineração

Para geração dos modelos, optou-se por utilizar nove algoritmos de mineração de dados diferentes. Sendo eles, Bayes Net (BN), Naive Bayes (NB), Naive Bayes Simple (NBS), Multilayer Perceptron (MP), RBF Network (RBF), Simple Logistic (SL), Decision Table (DT), Random Forest (RF) e J48 (J48). Onde cada um destes algoritmos possui suas particularidades e trabalha com diferentes métodos de classificação brevemente apresentados a seguir. A escolha destes algoritmos se deve a estes serem os mais utilizados em pesquisas relacionadas ao tema.

O classificador Bayes Net utiliza modelos probabilísticos, o Naive Bayes e o Naive Bayes Simple são classificadores probabilísticos baseados na aplicação do teorema de Bayes, o Multilayer Perceptron utiliza modelos de redes neurais artificiais, o RBF

Network utiliza redes neurais de aprendizagem não supervisionada, o Simple Logistic utiliza modelos de regressão lógica linear, o Decision Table utiliza tabelas de decisão e por último tanto o J48 quanto o Random Forest utilizam árvores de decisão.

5. Experimentos

Foram efetuados 3 experimentos, onde o primeiro e o segundo utilizam dados de polos de cidades diferentes e o terceiro é a junção dos dados dos dois primeiros. Esta separação tem como objetivo verificar se os polo são fatores desencadeantes do processo de evasão.

5.1. Experimento 1

Este experimento consistiu na execução dos nove algoritmos selecionados, no ambiente WEE configurando o mesmo para cross-validation 10 folds. Utilizou-se os dados da turma do Curso Técnico em Administração do Polo da primeira cidade, que era composto por 50 alunos, sendo que 40 concluíram o curso e outros 10 evadiram.

	Total de interações	Média de interações por aluno
Semana 1	2453	49,06
Semana 2	2701	54,02
Semana 3	864	17,28
Semana 4	3459	69,18
Total	9477	189,54

Tabela 1. Dados Polo Agudo

Como pode ser observado na tabela 1, o experimento contou com 9477 interações dos alunos divididas nas 4 semanas sendo a média de interações por aluno igual a 189,54. Os resultados obtidos após a ferramenta calcular a média de acurácia dos experimentos é apresentado na Tabela 2.

Algoritmo	Alcurácia (%)	Verd. Positivo (%)	Verd. Negativo (%)
BayesNet	80,00	100	0
NaiveBayes	62,00	67,50	40,00
NaiveBayes Simple	66,00	75,00	30
MultilayerPerceptron	80,00	100	0
Simple Logistic	80,00	100	0
DecisionTable	80,00	100	0
RandomForest	80,00	100	0
J48	78,00	92,50	20,00
RBF Network	80,00	87,50	50,00

Tabela 2. Resultados experimento 1

Observando os resultados pode se verificar que os algoritmos BayesNet, Multilayer Perceptron, Simple Logistic, Decision Table e RandomForest não obtiveram nenhum nível de aprendizagem, pois como o experimento tem 50 alunos no total e destes 10 são evadidos, teremos 80% de alunos concluintes e 20% de alunos evadidos. Assim

os resultados dos algoritmos citados, demonstram que eles apenas inferem que todos os alunos estão em situação normal, ou seja, não irão evadir.

Por outro lado, os algoritmos Naive Bayes, Naive Bayes Simple, J48 e RBF Network demonstraram algum nível de aprendizagem, onde o RBF Network obteve os resultados mais satisfatórios, tanto em acurácia onde obteve 80% quanto em predição de alunos evadidos onde obteve 50% de acertos.

5.2. Experimento 2

Neste experimento utilizou-se os dados da turma do Curso Técnico em Administração do Polo da segunda cidade, que é composto por 50 alunos. Entretanto destes, 16 acabaram as disciplinas iniciais sem interações e seu status final do curso consta como evadido. Como estes alunos tem um peso de 32% do conjunto de dados e são considerados de fácil predição, optou-se por excluí-los da classificação, pois entende-se que mantê-los no conjunto de dados acabaria mascarando os resultados. O restante da turma é composto por 34 alunos, onde destes 18 concluíram o curso e 16 evadiram, o que representa respectivamente 52,94% e 47,06% .

	Total de interações	Média de interações
Semana 1	686	20,17
Semana 2	1817	53,44
Semana 3	693	20,18
Semana 4	1263	37,14
Total	4459	131,14

Tabela 3. Dados Polo cidade 2

Como pode ser observado na tabela 3, este experimento contou com 4459 interações dos alunos nas 4 semanas, sendo que a média por aluno foi de 131,14 no total.

O ambiente utilizado para este experimento foi o WEE do WEKA configurado para cross-validation 10 folds e com os 9 algoritmos definidos anteriormente. Os resultados obtidos após o processamento dos dados são apresentados na tabela 4.

Algoritmo	Alcurácia (%)	Verd. Positivo (%)	Verd. Negativo (%)
BayesNet	67,64	83,33	50,00
NaiveBayes	61,76	55,55	68,75
NaiveBayes Simple	61,76	55,55	68,75
MultilayerPerceptron	61,76	55,55	68,75
Simple Logistic	76,46	72,22	81,25
DecisionTable	67,64	83,33	50,00
RandomForest	61,76	66,66	56,25
J48	67,64	77,77	56,25
RBF Network	50,01	50,00	50,00

Tabela 4. Resultados experimento 2

Analisando os resultados deste experimento, os mais satisfatórios em termos de acurácia geral foram obtidos pelo Simple Logistic, tendo em média 76,46% de precisão.

Entretanto, como o objetivo é a predição de alunos evadidos e não a acurácia geral, podemos observar que dentre o conjunto de algoritmos que foram testados, os que apresentaram um resultado mais satisfatório na predição de alunos evadidos foram Multilayer Perceptron, Bayesnet, Naive Bayes, Naive Bayes Simple e Simple Logistic. Este último se destaca pois dentre os 16 alunos evadidos do conjunto de dados ele conseguiu classificar 13 de forma correta, o que resulta em uma porcentagem de 81,25% de acurácia para evasão.

5.3. Experimento 3

O experimento 3 reuniu os dados das turmas do Curso Técnico em Administração dos Polos das cidades 1 e 2. Assim, ele é composto por 84 alunos, sendo que destes 58 concluíram o curso e 26 evadiram.

	Total de interações	Média de interações
Semana 1	3139	37,36
Semana 2	4518	53,78
Semana 3	1557	18,53
Semana 4	4722	56,21
Total	13936	165,90

Tabela 5. Dados Polos das cidades 1 e 2

Como pode ser observado na tabela 5, este experimento contou com 13.936 interações dos alunos nas 4 semanas, sendo que a média por aluno foi de 165,90 no total.

O ambiente WEE do WEKA foi utilizado neste experimento, com sua configuração definida para cross-validation 10 folds e com os 9 algoritmos definidos anteriormente. Os resultados obtidos após a execução do testes são apresentados na tabela 6.

Algoritmo	Acurácia (%)	Verd. Positivo (%)	Verd. Negativo (%)
BayesNet	79,76	82,75	73,07
NaiveBayes	73,80	75,86	69,23
NaiveBayes Simple	73,80	75,86	69,23
MultilayerPerceptron	71,42	91,37	26,92
Simple Logistic	71,42	86,20	38,46
DecisionTable	79,76	82,75	73,07
RandomForest	72,61	100	11,53
J48	76,19	81,03	65,38
RBF Network	69,04	79,31	46,15

Tabela 6. Resultados experimento 3

Observando os resultados deste experimento, podemos ver que as médias mais elevadas em acurácia geral foram obtidas pelos classificadores Bayes Net e pelo Decision Table, sendo estes também os classificadores a obterem a maior quantidade de acertos no objetivo principal deste projeto, predição precoce de evasão de alunos, tendo ambos 19 acertos dentro o conjunto 26 alunos em evasão, obtendo assim uma porcentagem de

73,07% na predição de alunos evadidos. Ainda vale ressaltar que alguns dos classificadores obtiveram resultados interessantes como o Naive Bayes, o Naive Bayes Simple e o J48, tendo obtido respectivamente 18 (69,23%), 18 (69,23%), 17 (65,38%).

5.4. Discussão dos Resultados

Analisando os resultados de verdadeiros negativos obtidos e tendo em vista que o objetivo do trabalho é a predição precoce de alunos em risco de evasão, podemos notar que o balanceamento dos dados é de suma importância para a aprendizagem dos algoritmos. Assim quanto mais balanceado os dados de treinamento forem, melhores os resultados obtidos nos testes.

No segundo experimento o algoritmo que apresentou o resultado mais satisfatório em verdadeiro negativo foi o Simple Logistic com 81,25%, sendo esta ainda a taxa mais alta de acertos obtida em todos os experimentos. Enquanto que os algoritmos com piores resultados foram o Bayes Net e o RBF Network ambos com 50%. Se compararmos os resultados dos dois primeiros experimentos, os piores do segundo ainda são no mínimo iguais aos melhores do primeiro, no caso o RBF Network com 50%.

No terceiro experimento os resultados obtidos em acurácia geral foram os mais satisfatórios, ignorando os resultados dos algoritmos que não obtiveram nenhuma aprendizagem do primeiro experimento, tanto o Bayes Net quanto o Decision Table apresentando 79,76% de acurácia geral, sendo estes dois algoritmos também os que obtiveram as maiores médias em verdadeiro negativo no experimento com 73,07% ambos.

Observando os resultados do primeiro experimento, podemos notar que apenas 4 dos algoritmos testados obtiveram algum nível real de aprendizagem, mas este foi o conjunto de dados mais desbalanceados dos testes, sendo apenas 20% de evadidos do total de 50 alunos. Isso auxilia a demonstrar que o balanceamento dos dados é essencial nas etapas de treinamento.

6. Considerações Finais e Trabalhos Futuros

Como os resultados dos experimentos demonstram, a predição precoce de alunos em risco de evasão utilizando técnicas de mineração de dados nos log's das 4 primeiras semanas do curso é viável. Assim, tornando possível o desenvolvimento de uma ferramenta de auxílio acadêmico que identifique os alunos que tem risco de evasão precocemente e que permita que estes alunos tenham um tratamento diferenciado pelos responsáveis, evitando a evasão destes alunos.

Entretanto, vários dos algoritmos testados acabaram não tendo nenhum nível de aprendizagem no experimento 1, onde os dados eram pouco balanceados, sendo necessário a coleta de uma quantidade maior de dados e testes com outros algoritmos e configurações a obtenção de resultados mais significativos em todos os casos.

Como o presente trabalho ainda encontra-se em fase inicial, em sua evolução a coleta, pré-processamento e classificação de uma quantidade maior de dados de outras turmas é necessária, bem como a aplicação de procedimentos semelhantes nos outros 4 cursos ofertados pela instituição.

Como trabalho futuro, fica tanto a possibilidade de se trabalhar com a contagem dos diferentes tipo de interações e não apenas com as interações em sua totalidade, como

a possibilidade de alterar o número de semanas utilizadas. Assim buscando melhorar a classificação precoce de alunos em situação de evasão.

Referências

- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.
- Barroso, M. F. and Falcão, E. B. (2004). Evasão universitária: O caso do instituto de física da ufrj. *Encontro Nacional de Pesquisa em Ensino de Física*, 9:1–14.
- Bouckaert, R. Eibe, F. M. H. a. K. R. a. R. P. a. S. A. a. S. D. (2010). Weka manual for version 3-6-4. *University of Waikato*.
- Cambruzzi, W. L. (2014). GVWISE: Uma aplicação de learning analytics para a redução da evasão na educação a distância.
- Delano, R. and Corrêa, D. S. (2013). Redes na educação a distância: Uma análise estrutural do sistema uab em minas gerais.
- Detoni, D., Araujo, R. M., and Cechinel, C. (2014). Predição de reprovação de alunos de educação a distância utilizando contagem de interações. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 25, pages 896–905.
- 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 *Anais do Simpósio Brasileiro de Informática na Educação*, volume 24.
- 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*.
- Rigo, S. (2014). Evasão na ead - do diagnóstico à estratégia: uma metodologia sistêmica utilizando cartografia dos fluxos, mineração de dados e learning analytics. *Anais do 10º seminário nacional de educação a Distância*.
- Segundo, F. R. and Ramos, D. K. (2005). Soluções baseadas no uso de software livre: alternativas de suporte tecnológico à educação presencial e a distância. *Anais do 12 Congresso Internacional de Educação a Distância*, 12:18–22.