

UNIVERSIDADE DE PERNAMBUCO
CAMPUS GARANHUNS

ROZELMA SOARES DE FRANÇA

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS, BASEADA NA
TAXONOMIA DE BLOOM, PARA O MAPEAMENTO DO CONHECIMENTO NA
APRENDIZAGEM DE PROGRAMAÇÃO**

GARANHUNS
JANEIRO, 2013

ROZELMA SOARES DE FRANÇA

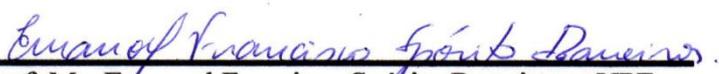
**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS, BASEADA NA
TAXONOMIA DE BLOOM, PARA O MAPEAMENTO DO CONHECIMENTO NA
APRENDIZAGEM DE PROGRAMAÇÃO**

Monografia apresentada como requisito parcial para
obtenção do diploma de Licenciado em Computação
pela Universidade de Pernambuco - *Campus*
Garanhuns.

Orientador: Prof. Ms. Haroldo José Costa do Amaral

**GARANHUNS
JANEIRO, 2013**

Monografia de Graduação apresentada por **Rozelma Soares de França** do Curso de Graduação em Licenciatura em Computação da Universidade de Pernambuco, *Campus* Garanhuns, sob o título “**Aplicação de Técnicas de Mineração de Dados, Baseada na Taxonomia de Bloom, para o Mapeamento do Conhecimento na Aprendizagem de Programação**”, orientada pelo **Prof. Ms. Haroldo José Costa do Amaral** e aprovada pela Banca Examinadora formada pelos professores:



Prof. Ms. Emanuel Francisco Sposito Barreiros – UPE
Avaliador



Prof. Ms. Haroldo José Costa do Amaral – UPE
Orientador

Visto e permitida à impressão.
Garanhuns, 07 de janeiro de 2013.



Prof. Ms. Haroldo José Costa do Amaral
Coordenador do Curso de Licenciatura em Computação
da Universidade de Pernambuco, *Campus* Garanhuns.

Ao meu pai (*in memoriam*), João Soares da Silva,
meu amor e gratidão. Eternas saudades!

AGRADECIMENTOS

Ao meu orientador Prof. Haroldo José Costa do Amaral minha gratidão pelas oportunidades concedidas, orientação, apoio e amizade.

Ao membro da banca, Prof. Emanuel Francisco Spósito Barreiros, pela disposição e interesse em contribuir para o melhoramento deste trabalho.

Aos professores que tive durante minha vida acadêmica que de alguma forma contribuíram para minha formação.

À minha mãe, pelo amor e amizade e por sempre apostar e acreditar em meus ideais.

Aos meus irmãos, Reginaldo, Roberto e Robson, pelo carinho, confiança e apoio nas minhas decisões.

Aos meus sobrinhos, Bianca, Ítalo, Kayky, João Vitor, Roberta e Robson Junior, pela ternura e doçura constantes.

Aos meus amigos pelo apoio e incentivo para conclusão deste trabalho. Em especial, meus agradecimentos a Adgina, Aldilane, Cláudia, Gelmire, Katharine, Jullyane, Paulo e Waldir.

Por fim, meu agradecimento maior a Deus, fonte de amor e vida.

RESUMO

O processo contínuo de avaliação da aprendizagem pode gerar um grande volume de dados inviável de ser manipulado pelo professor. Tais dados podem esconder informações úteis acerca do nível cognitivo do aprendiz e de suas necessidades de aprendizagem. A automatização do tratamento e análise dos dados pode ajudar nessa tarefa e propiciar a extração de informações que auxiliem professores na gestão do processo de ensino. Neste trabalho, propõe-se a definição de um conjunto de atributos, pautado na Taxonomia de Bloom, para aplicação de técnicas de mineração de dados, que possibilite realizar o mapeamento do conhecimento de um grupo de aprendizes, em dados coletados através de avaliações. Com o objetivo de verificar a adequação do conjunto de atributos proposto, realizaram-se experimentos em dados provenientes de avaliações formativas de uma disciplina de programação. Espera-se, com isso, permitir o diagnóstico precoce de dificuldades na aprendizagem e apoiar a realização de ações pedagógicas que favoreçam o processo de ensino-aprendizagem nas modalidades de ensino presencial e à distância.

Palavras-chave: Mineração de dados. Avaliação formativa. Taxonomia de Bloom.

ABSTRACT

The continuous process of learning assessment may generate a large volume of data impossible to be manipulated by the teacher. Such data can hide useful information about the cognitive level of the learner and their learning needs. Automating the processing and analysis of data can help in this task and provide the extraction of information to help teachers in the management of the teaching process. In this work, we propose the definition of a set of attributes, based on Bloom's Taxonomy for application of data mining techniques, enabling realize the knowledge mapping of a group of learners on data collected through assessments. Aiming to verify the adequacy of the proposed set of attributes, experiments were conducted on data from formative assessments of a programming discipline. It is hoped, therefore, allow early diagnosis of learning difficulties and support the realization of educational activities that promote the teaching-learning process in terms of classroom teaching and distance learning.

Keywords: Data Mining. Formative assessment. Bloom's Taxonomy.

LISTA DE FIGURAS

Figura 1 – Categorias do domínio cognitivo da Taxonomia de Bloom	14
Figura 2 – Categorias do domínio cognitivo da Taxonomia de Bloom revisada	15
Figura 3 – Código da questão sobre <i>array</i> / Categoria Analisar.	18
Figura 4 – Exemplo Árvore de Decisão	22
Figura 5 – Saída do algoritmo J4.8 com atributo alvo <i>ConceitoAvaliacao</i>	30
Figura 6 – Árvore de decisão gerada com atributo alvo <i>ConceitoAvaliacao</i>	32
Figura 7 – Saída do algoritmo J4.8 com atributo alvo <i>RotuloSomativa</i>	34
Figura 8 – Árvore de decisão gerada com atributo alvo <i>RotuloSomativa</i>	35
Figura 9 – Regras de associação geradas pelo algoritmo <i>Apriori</i>	36
Figura 10 – Agrupamento em <i>clusters</i>	37

LISTA DE QUADROS

Quadro 1 – Estrutura do processo cognitivo na Taxonomia de Bloom revisada	16
Quadro 2 – Fases e Etapas do KDD	20
Quadro 3 – Classificação das questões das avaliações formativas de acordo com a Taxonomia de Bloom	27
Quadro 4 – Sequência temporal das avaliações formativas	28
Quadro 5 – Atributos propostos para o mapeamento do conhecimento do estudante.....	28

LISTA DE ABREVIATURAS

AVA	Ambientes Virtuais de Aprendizagem
DM.....	<i>Data Mining</i>
EDM	<i>Educational Data Mining</i>
KDD	<i>Knowledge Discovery in Database</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

SUMÁRIO

1. INTRODUÇÃO	11
2. A TAXONOMIA DE BLOOM	13
2.1. A TAXONOMIA DE BLOOM REVISADA	14
2.2. A TAXONOMIA DE BLOOM NO CONTEXTO DA COMPUTAÇÃO	16
2.2.1. CATEGORIA <i>LEMBRAR</i>	17
2.2.2. CATEGORIA <i>ENTENDER</i>	17
2.2.3. CATEGORIA <i>APLICAR</i>	17
2.2.4. CATEGORIA <i>ANALISAR</i>	18
2.2.5. CATEGORIA <i>AVALIAR</i>	18
2.2.6. CATEGORIA <i>CRIAR</i>	19
3. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS	20
3.1. FASES DO KDD	21
3.2. TAREFAS E TÉCNICAS DE MINERAÇÃO DE DADOS	21
3.2.1. CLASSIFICAÇÃO	21
3.2.2. ASSOCIAÇÃO	22
3.2.3. AGRUPAMENTO	23
3.3. MINERAÇÃO DE DADOS NA EDUCAÇÃO	23
4. ESTUDO DE CASO	26
5. EXTRAÇÃO DE PADRÕES E ANÁLISE DE RESULTADOS	30
5.1. TAREFA DE CLASSIFICAÇÃO	30
5.2. TAREFA DE ASSOCIAÇÃO	36
5.3. TAREFA DE AGRUPAMENTO	37
6. CONSIDERAÇÕES FINAIS	40
REFERÊNCIAS BIBLIOGRÁFICAS	42

1. INTRODUÇÃO

A avaliação é essencial à educação e indissociável enquanto concebida como problematização, questionamento e reflexão sobre a ação (HOFFMANN, 2010). Nesta perspectiva, ela desempenha um papel fundamental em promover a aprendizagem, produzindo informações que poderão ajudar estudantes e professores. A avaliação, neste paradigma, deixa de ser apenas um instrumento de verificação da aprendizagem e passa atuar diretamente no processo de ensino-aprendizagem permeando e auxiliando todo este processo, não mais como uma atividade em momentos estanques e pontuais (CERNY, 2001).

Há uma busca por uma avaliação formativa que pode ser entendida como uma prática de avaliação contínua cujo objetivo principal é melhorar as aprendizagens em curso, contribuindo para o acompanhamento e orientação dos estudantes durante todo seu processo de formação (PERRENOUD, 1999). Esta modalidade de avaliação, segundo Hadji (2001), possui características informativa e reguladora, ou seja, fornece informações aos dois atores do processo de ensino-aprendizagem: i) professor, que será informado dos efeitos reais de suas ações, podendo regular sua ação pedagógica e ii) aprendiz, que terá a oportunidade de tomar consciência de suas dificuldades e, possivelmente, reconhecer e corrigir seus erros.

O processo de avaliação contínua da aprendizagem pode produzir uma grande massa de dados no decorrer do tempo requerendo procedimentos automáticos ou semiautomáticos para tratamento e análise destes dados. Justifica-se, portanto, a pesquisa para a aplicação de técnicas computacionais que possam dar suporte eficiente a esta atividade. Nesse contexto, a comunidade de pesquisa em Informática na Educação tem investigado possibilidades de exploração dessas informações através da utilização de técnicas de mineração de dados. Espera-se, com isso, que as possíveis descobertas realizadas tenham o potencial de auxiliar na compreensão de como os estudantes aprendem e identificar os principais fatores que causam impactos no processo de aprendizagem (GOTTARDO et al., 2012).

Pesquisadores de Mineração de Dados Educacionais – *Educational Data Mining (EDM)* - citam algumas possibilidades de pesquisa na área, a exemplo de: i) modelos para identificar alunos com dificuldades de aprendizagem e ii) meios de melhorar a qualidade do material didático (BAKER et al., 2011).

Em se tratando da aplicação de técnicas de mineração em dados obtidos a partir de avaliações da aprendizagem, pode-se mencionar as pesquisas de Pimentel et al. (2006) e Santos et al. (2012). Ressalta-se, no entanto, que nos estudos desses autores, não se observa, durante a seleção dos dados, uma das etapas do KDD (*Knowledge Discovery in Database*), a inclusão de atributos que façam referência aos objetivos pedagógicos planejados e que se deseja serem alcançados pelos aprendizes. Assim, este trabalho tem como objetivo apresentar um conjunto de atributos elaborado a partir de uma taxonomia de objetivos educacionais, a Taxonomia de Bloom (BLOOM et al., 1956; ANDERSON & KRATHWOHL, 2001), para aplicação de técnicas de mineração de dados em avaliações contínuas da aprendizagem, que possibilite realizar o mapeamento do conhecimento dos estudantes. Com isso, será possível identificar quais conceitos e em que níveis eles foram assimilados pelos aprendizes possibilitando, assim, que estratégias pedagógicas sejam adequadamente planejadas para grupos de aprendizes que apresentem as dificuldades identificadas.

Para validar a proposta apresentada, um estudo de caso foi realizado envolvendo a disciplina Programação Orientada a Objetos do curso de Licenciatura em Computação da Universidade de Pernambuco. Esta escolha deu-se pelo fato de que tal disciplina tem apresentado altas taxas de reprovação, alcançando em 2010 uma taxa superior a 75%. Tal realidade não é diferente em outros cursos de graduação, na área de Computação, sendo considerado um dos gargalos existentes, dificultando ou até mesmo impedindo a continuidade dos alunos em seus estudos (RAPKIEWICZ et al., 2006).

O restante do trabalho está organizado como segue: a Seção 2 apresenta uma taxonomia de objetivos educacionais, tida como referencial pedagógico para a elaboração do conjunto de atributos proposto e a aplicação desta taxonomia do contexto da Computação. A Seção 3 apresenta conceitos sobre mineração de dados e algoritmos para extração de padrões e traz alguns trabalhos que aplicaram técnicas de mineração de dados no âmbito da educação. Nas Seções 4 e 5, descreve-se, respectivamente, um estudo de caso aplicando técnicas de mineração em dados obtidos em sessões de avaliação da aprendizagem e os resultados obtidos a partir de tal tarefa. Já na Seção 6 são feitas algumas considerações acerca deste trabalho e os aprofundamentos necessários.

2. A TAXONOMIA DE BLOOM

Na educação, decidir e definir os objetivos de aprendizagem significa estruturar o processo educacional de modo a oportunizar mudanças de pensamentos, ações e condutas, de forma consciente (FERRAZ et al., 2010). Quando os objetivos estão bem definidos, pode ficar mais fácil para os discentes atingirem o nível de desenvolvimento cognitivo, já que saberão exatamente o que deles é esperado durante e após o processo de ensino.

Segundo Ferraz et al. (2010), a clara e estruturada definição de objetivos educacionais direcionará o processo de ensino para a escolha adequada de estratégias, métodos, delimitação do conteúdo específico, instrumentos de avaliação e, conseqüentemente, para uma aprendizagem efetiva e duradoura. Neste contexto, um dos instrumentos existentes que pode vir a facilitar esse processo é a taxonomia proposta por Bloom et al. (1956), conhecida como Taxonomia de Bloom, que tem como objetivo, ajudar no planejamento, organização e controle dos objetivos de aprendizagem.

A Taxonomia de Bloom é composta por três domínios:

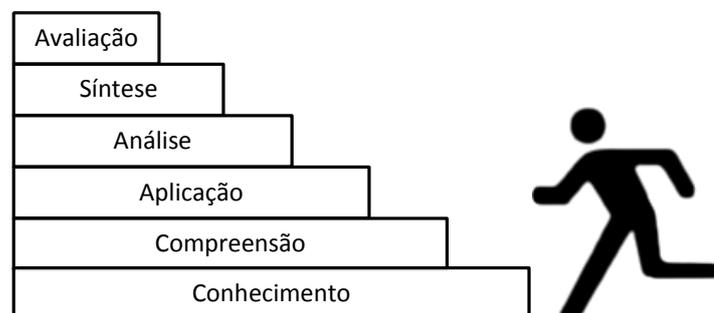
- **Cognitivo:** relaciona-se com o aprender, dominar um conhecimento. Envolve a aquisição de um novo conhecimento, do desenvolvimento intelectual, de habilidade e de atitudes. Nesse domínio, os objetivos foram agrupados em seis diferentes categorias e são apresentados numa hierarquia de complexidade e dependência, do mais simples ao mais complexo. Assim, para ascender a uma nova categoria, é preciso ter obtido um adequado desempenho na anterior, pois cada uma utiliza capacidades adquiridas nos níveis anteriores. Como pode ser visto na Figura 1, as categorias desse domínio são: Conhecimento; Compreensão; Aplicação; Análise; Síntese; e Avaliação;
- **Afetivo:** relacionado a sentimentos e posturas. Envolve categorias ligadas ao desenvolvimento da área emocional e afetiva, que incluem comportamento, atitude, responsabilidade, respeito, emoção e valores. Para ascender a uma nova categoria é preciso ter obtido um desempenho adequado na anterior, pois cada uma utiliza capacidades adquiridas nos níveis anteriores para serem aprimoradas. As categorias desse domínio são: Receptividade; Resposta; Valorização; Organização; e Caracterização;
- **Psicomotor:** associado a habilidades físicas específicas. Apesar de Bloom e seus colaboradores não terem definido uma taxonomia para a área psicomotora, outros

pesquisadores o fizeram, segundo Ferraz et al. (2010), e chegaram a seis categorias que incluem ideias ligadas a reflexos, percepção, habilidades físicas, movimentos aperfeiçoados e comunicação não verbal. Para ascender a uma nova categoria, é preciso ter obtido um desempenho adequado na anterior, pois cada uma utiliza capacidades adquiridas nos níveis anteriores. As categorias desse domínio são: Imitação; Manipulação; Articulação; e Naturalização.

Todos os três níveis foram amplamente discutidos e divulgados, em momentos distintos e por pesquisadores diferentes, entretanto, o domínio cognitivo é o mais conhecido e utilizado. Muitos educadores se apoiam nos pressupostos teóricos desse domínio para definirem, em seus planejamentos educacionais, objetivos, estratégias e sistemas de avaliação (FERRAZ et al., 2010).

Relacionados às categorias, estão os verbos que buscam dar suporte ao planejamento acadêmico (objetivo, estratégia e avaliação) relacionado a cada uma delas.

Figura 1 – Categorias do domínio cognitivo da Taxonomia de Bloom



Fonte: Elaborada pelo autor

2.1. A TAXONOMIA DE BLOOM REVISADA

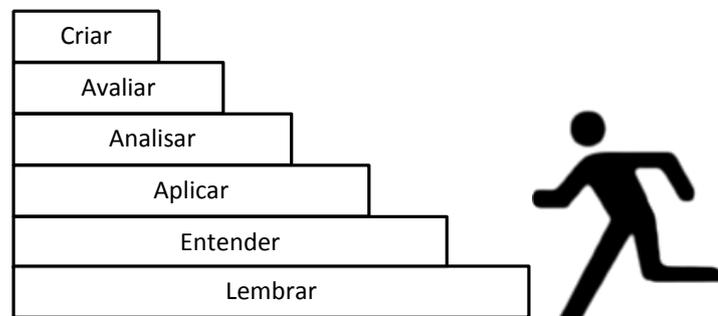
Com a incorporação de novos conceitos, recursos e teorias ao campo educacional, bem como os avanços psico-pedagógicos e tecnológicos ocorridos e a publicação de diversas experiências de sucesso no uso efetivo da taxonomia, observou-se a necessidade de reavaliação e releitura dos pressupostos teóricos que sustentam a pesquisa original apresentada em 1956 (BLOOM et al., 1956) e, assim, foram realizadas revisão e atualização da Taxonomia de Bloom (ANDERSON & KRATHWOHL, 2001).

Na taxonomia revisada, a base das categorias foi mantida, permanecendo seis categorias, entretanto, ao separar, conceitualmente, o conhecimento do processo cognitivo, ocorreram as seguintes mudanças (KRATHWOHL, 2002):

- Os aspectos verbais utilizados na categoria Conhecimento foram mantidos, mas esta foi renomeada para Lembrar;
- Compreensão foi renomeada para Entender;
- Aplicação, Análise e Avaliação, foram alteradas para a forma verbal Aplicar, Analisar e Avaliar, respectivamente, por expressarem melhor a ação pretendida e serem condizentes com o que se espera de resultado a determinado estímulo de instrução;
- Síntese trocou de lugar com Avaliação (Avaliar) e foi rebatizada para a forma verbal Criar;
- Os nomes das subcategorias existentes foram alterados para verbos no gerúndio (Quadro 1).

Na Figura 2, encontra-se a categorização atual da Taxonomia de Bloom.

Figura 2 – Categorias do domínio cognitivo da Taxonomia de Bloom revisada



Fonte: Elaborada pelo autor

Apesar de manter o *design* hierárquico da original, a nova taxonomia é flexível, pois, como aponta Krathwohl (2002), considerou a possibilidade de interpolação das categorias do processo cognitivo quando necessário, devido ao fato de que determinados conteúdos podem ser mais fáceis de serem assimilados a partir do estímulo pertencente a uma mais complexa. Por exemplo, pode ser mais fácil entender um assunto após aplicá-lo e só então ser capaz de explicá-lo.

Quadro 1 – Estrutura do processo cognitivo na Taxonomia de Bloom revisada

Lembrar
Relacionado a reconhecer e reproduzir ideias e conteúdos. Reconhecer requer distinguir e selecionar uma determinada informação e reproduzir ou recordar está mais relacionado à busca por uma informação relevante memorizada. Representado pelos seguintes verbos no gerúndio: Reconhecendo e Reproduzindo.
Entender
Relacionado a estabelecer uma conexão entre o novo e o conhecimento previamente adquirido. A informação é entendida quando o aprendiz consegue reproduzi-la com suas “próprias palavras”. Representado pelos seguintes verbos no gerúndio: Interpretando, Exemplificando, Classificando, Resumindo, Inferindo, Comparando e Explicando.
Aplicar
Relacionado a executar ou usar um procedimento numa situação específica e pode também abordar a aplicação de um conhecimento numa situação nova. Representado pelos seguintes verbos no gerúndio: Executando e Implementando.
Analisar
Relacionado a dividir a informação em partes relevantes e irrelevantes, importantes e menos importantes e entender a inter-relação existente entre as partes. Representado pelos seguintes verbos no gerúndio: Diferenciando, Organizando, Atribuindo e Concluindo.
Avaliar
Relacionado a realizar julgamentos baseados em critérios e padrões qualitativos e quantitativos ou de eficiência e eficácia. Representado pelos seguintes verbos no gerúndio: Checando e Criticando.
Criar
Significa colocar elementos junto com o objetivo de criar uma nova visão, uma nova solução, estrutura ou modelo utilizando conhecimentos e habilidades previamente adquiridos. Envolve o desenvolvimento de ideias novas e originais, produtos e métodos por meio da percepção da interdisciplinaridade e da interdependência de conceitos. Representado pelos seguintes verbos no gerúndio: Generalizando, Planejando e Produzindo.

Fonte: Ferraz et al. (2010)

2.2. A TAXONOMIA DE BLOOM NO CONTEXTO DA COMPUTAÇÃO

A taxonomia de Bloom é bastante conhecida e utilizada em várias áreas do conhecimento, inclusive na Computação. Pesquisas demonstram a aplicabilidade dessa taxonomia na elaboração de instrumentos de avaliação na disciplina de programação, contudo, segundo Whalley et al. (2006), muitas das descrições dos níveis da taxonomia são difíceis de serem interpretadas no contexto dos exercícios dessa disciplina. Por conta deste tipo de dificuldade, a seguir, descreve-se cada uma das categorias da taxonomia de Bloom e exemplos de como essas categorias podem ser interpretadas e utilizadas no contexto da Computação.

2.2.1. CATEGORIA *LEMBRAR*

Essa categoria requer que o aluno recupere informações da memória. Nesse sentido, no contexto da Computação, Scott (2003) cita três exemplos de tarefas que podem ser classificadas no nível Lembrar da taxonomia: 1) Citar os nomes dos três tipos de *loops* em C++; 2) Listar três operações de entrada/saída em um computador; e 3) Escrever cinco coisas que são verdadeiras sobre uma arquitetura RISC.

2.2.2. CATEGORIA *ENTENDER*

Nesta categoria, é requerido do aprendiz que dê um significado ao material ou experiências educacionais. Interpretar, Exemplificar, Classificar, Resumir, Concluir, Comparar, Explicar são verbos que podem ser utilizados em instrumentos de avaliação baseados nesse nível cognitivo (KRATHWOHL, 2002).

Jesus et al. (2009) exemplificam o nível Entender através de algumas tarefas: 1) os estudantes recebem um trecho de código que apenas exibe o valor do contador de um laço de repetição. Em seguida, são apresentados 4 fluxogramas. O aprendiz deve dizer qual dos fluxogramas possui a mesma lógica do trecho de código apresentado; 2) os alunos recebem um diagrama e quatro trechos de código semelhantes ao apresentado na tarefa descrita anteriormente. Depois, eles devem indicar qual dos trechos de código possui a mesma lógica do diagrama. Nesta questão, o aluno faz o processo inverso daquele que é feito na questão anterior.

2.2.3. CATEGORIA *APLICAR*

Jesus et al. (2009) exemplificam esta categoria com uma atividade em que os estudantes recebem um trecho de código onde um valor é solicitado ao usuário e em seguida dois desvios condicionais fazem testes lógicos com o valor solicitado. Um dos desvios é aninhado no outro. O primeiro desvio condicional verifica se o valor solicitado é par e o segundo, o aninhado, verifica se o mesmo valor é positivo. Dentro do desvio condicional mais interno existe um comando que exibe uma mensagem que está omitida, pois os alunos devem indicar qual dentre

4 mensagens é a correta para a situação. As opções possíveis são: a) “par e positivo”; b) “ímpar e positivo”; c) “maior que dois”; e d) “positivo ou zero”.

2.2.4. CATEGORIA ANALISAR

Verbos como Diferenciar, Organizar, Atribuir estão associados a esta categoria (KRATHWOHL, 2002). Scott (2003) fornece um exemplo que pode ser classificado no nível Analisar: Comparar as vantagens e desvantagens entre as arquiteturas RISC e CISC. Jesus et al. (2009) propõem uma atividade em que os estudantes recebem um trecho de código que ordena um *array* em ordem ascendente. O algoritmo está incompleto, como mostra a Figura 3, e o aluno deve indicar a sequência de instruções que completa o algoritmo corretamente. Para isso, o aluno deve ser capaz de indicar qual dentre as opções seguintes é a sequência de instruções correta.

- a) 4, 2, 6, 3, 1, 5 b) 2, 4, 6, 5, 1, 3 c) 4, 2, 6, 5, 1, 3 d) 2, 4, 6, 1, 5, 3

Figura 3 – Código da questão sobre *array* / Categoria Analisar.

<pre> programa ex11 declarações defina TAM 5 inteiro temp, i, j inteiro array[TAM] inicio array <- {4, 3, 2, 1, 0} ??? ??? ??? ??? fimse fimpara fimpara fim </pre>	<p>Instruções que completam o algoritmo:</p> <ol style="list-style-type: none"> 1) array[i] <- array[j] 2) para j <- i ate TAM-1 passo 1 3) array[j] <- temp 4) para i <- 0 ate TAM-1 passo 1 5) temp <- array[i] 6) se (array[j] < array[i]) entao
--	---

Fonte: Jesus et al. (2009)

2.2.5. CATEGORIA AVALIAR

Scott (2003) cita alguns exemplos de questões que se enquadram na categoria Avaliar. O autor propõe que os alunos: 1) encontrem um erro de lógica presente em um trecho de código dado; 2) forneçam conjuntos de dados de teste para verificar um código e expliquem o que está

sendo testado em cada caso; 3) critiquem um código dado levantando seus pontos positivos e negativos.

2.2.6. CATEGORIA *CRIAR*

Os exemplos desta categoria fornecidos por Thompson et al. (2008) são: 1) propor um novo algoritmo alternativo ou hipotetizar que uma nova combinação de algoritmos resolverá o problema; 2) conceber um processo alternativo ou estratégia para resolver um problema; e 3) construir um segmento de código ou programa utilizando algoritmos inventados ou aplicando algoritmos conhecidos em uma combinação ainda não utilizada pelo aluno.

3. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

Com o avanço das tecnologias computacionais que permitem o armazenamento e processamento de um grande volume de dados, novas tecnologias têm sido desenvolvidas para auxiliar a extração de informações dessas bases de dados, destacando-se o *Knowledge Discovery in Database* (KDD) e o *Data Mining* (DM).

Segundo Fayyad (1996), KDD consiste num processo de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados. Ele é um processo abrangente de descoberta de conhecimento útil, a partir de dados, enquanto mineração de dados restringe-se à aplicação de algoritmos de extração de padrões (FAYYAD et al., 2002). Este trabalho debruça-se sobre a aplicação de KDD na área de Educação, na tentativa de descobrir padrões em dados de avaliações da aprendizagem, focando-se na etapa de Mineração de Dados.

O processo de descoberta do conhecimento envolve várias fases e etapas. Neste trabalho, o processo de KDD é classificado em três grandes fases (Quadro 2), tendo em vista as abordagens descritas em Rezende et al., (2003) e Pimentel et al. (2006): Preparação, Extração de Padrões e Pós-Processamento. A descrição de cada fase é apresentada na subseção a seguir, tendo como referencial os autores supramencionados.

Quadro 2 – Fases e Etapas do KDD

FASE	ETAPAS		OBJETIVO
Preparação	1	Seleção dos Dados	Compreender o domínio da aplicação, estabelecer objetivos e definir a massa de dados inicial a ser utilizada.
	2	Pré-processamento e Limpeza	Eliminar eventuais inconsistências, incompletudes, problemas com tipos de dados, etc.
	3	Transformação dos Dados	Transformar os dados para uma forma mais apropriada para a mineração, assumindo um formato estatístico e menos transacional.
Extração de Padrões	4	Mineração dos Dados	Aplicar métodos (algoritmos) para a Extração de Padrões.
Pós-Processamento	5	Avaliação dos Padrões	Analisar os resultados da etapa anterior a fim de identificar conhecimento nos padrões encontrados.
	6	Apresentação do Conhecimento	Utilizar técnicas de visualização e representação do conhecimento para apresentar ao usuário o conhecimento minerado.

Fonte: Pimentel et al. (2006)

3.1. FASES DO KDD

Na fase de **Preparação**, ocorre a identificação e o entendimento do problema, considerando aspectos como objetivos e as fontes de dados da qual se pretende extrair o conhecimento. O passo seguinte consiste na seleção dos dados a partir das fontes, de acordo com os objetivos do processo, e tratamento dos dados para poderem ser submetidos aos métodos e ferramentas na etapa de extração de padrões.

Na fase de **Extração de Padrões**, escolhe-se inicialmente um ou mais algoritmos para a extração de conhecimentos. Esta fase pode ser executada diversas vezes para o ajuste de parâmetros de maneira a se obter resultados adequados aos objetivos estabelecidos.

Na fase de **Pós-Processamento**, o conhecimento extraído é avaliado quanto a sua qualidade e/ou utilidade para que, em caso positivo, possa ser utilizado para apoiar algum processo de tomada de decisão, seja por um especialista humano ou por um sistema especialista.

3.2. TAREFAS E TÉCNICAS DE MINERAÇÃO DE DADOS

No processo de KDD, a mineração de dados é a etapa central que executa a descoberta de conhecimento propriamente dita, pois são os algoritmos de mineração de dados que produzem, de maneira semi-automática, o conhecimento a partir de dados existentes.

Segundo Pimentel et al. (2006), as tarefas de mineração de dados são as classes de problemas determinadas de acordo com o tipo de conhecimento a ser minerado e também dos objetivos almejavéis para a solução. A escolha da técnica de mineração e do algoritmo a ser utilizado, contudo, depende da tarefa que se quer executar.

Descrevem-se, nas subseções a seguir, as tarefas e técnicas de mineração de dados que foram utilizadas no estudo de caso descrito nas Seções 4 e 5 deste trabalho.

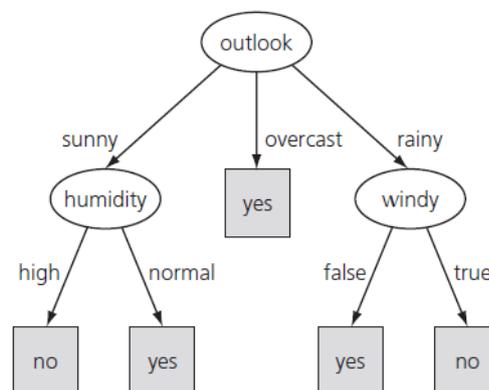
3.2.1. CLASSIFICAÇÃO

Conforme Han & Kamber (2001), classificação de dados consiste no processo de prever a classe de um objeto segundo certos atributos de predição.

Para construção do modelo de classificação, deve-se descobrir uma relação entre os atributos de predição e o atributo de classificação. Para tal, o conjunto de dados é fragmentado em duas amostras denominadas de conjunto de treinamento e conjunto de teste. O primeiro conjunto é usado para que o algoritmo de classificação descubra a relação anteriormente descrita. O segundo conjunto é utilizado para verificar a precisão do classificador gerado. O modelo de classificação resultante é então usado para prever a classe de registros não classificados (CHEN et al., 1996).

Uma das técnicas utilizadas para a tarefa de classificação é *Árvore de Decisão (Decision Tree)* que consiste em se escolher uma variável que se quer avaliar e o algoritmo procura as características mais fortemente relacionadas a ela, montando uma árvore de ramificações na forma SE-ENTÃO, conforme Figura 4. O J4.8 é um dos algoritmos que pode implementar esta técnica (WITTEN et al., 2011). Este algoritmo é uma implementação do algoritmo C.45 proposto por Quinlan (1993), para geração *top-down* de árvores de decisão. Segundo, Witten et al. (2011), o J4.8 utiliza uma técnica de busca gulosa, determinando a cada passo qual o atributo, dentre os disponíveis, é mais preditivo, e dividindo um nodo da árvore com base neste atributo.

Figura 4 – Exemplo *Árvore de Decisão*



Fonte: Witten & Frank (2005)

3.2.2. ASSOCIAÇÃO

Segundo Carvalho (2000), uma regra de associação é definida como: *se X então Y* ou $X \rightarrow Y$, onde X e Y são conjuntos de itens no formato *atributo-valor* e $X \cap Y = \Phi$. Diz-se que X é o antecedente da regra e Y é o conseqüente da regra. A meta é descobrir se X implica Y.

A técnica de regra de associação implementa este tipo de tarefa através de algoritmos como o *Apriori*. Segundo Witten & Frank (2005), este algoritmo, na ferramenta WEKA, possui as seguintes características: o valor de suporte mínimo começa com 100% e vai reduzindo 5% até que 10 regras sejam geradas com confiança mínima de 90% ou até que o valor de suporte atinja um limite inferior a 10%. Esses valores padrões podem ser alterados.

3.2.3. AGRUPAMENTO

A tarefa de agrupamento visa agrupar um conjunto de dados utilizando algum critério de similaridade e, diferentemente da tarefa de classificação, estes dados não devem ter sido previamente classificados (LOPES, 2003).

A técnica de particionamento implementa este tipo de tarefa sendo o algoritmo *K-means*, o mais conhecido (WITTEN et al., 2011). Este algoritmo, basicamente, divide o conjunto de dados em grupos (*clusters*), sendo necessário definir inicialmente o número de *clusters* que serão criados. Esse número é chamado de K e por isto o nome *K-means*.

3.3. MINERAÇÃO DE DADOS NA EDUCAÇÃO

Técnicas de mineração de dados podem ser aplicadas a uma variedade de contextos de tomada de decisão como *marketing*, finanças, manufatura, saúde. Analogamente, é possível minerar dados de alunos para verificar a relação entre uma abordagem pedagógica e o seu aprendizado e, através desta informação, o professor poder compreender se sua abordagem está realmente ajudando o estudante e desenvolver métodos de ensino mais eficazes (BAKER et al., 2011).

Em sistemas educacionais, o conhecimento descoberto pode ser útil tanto para os educadores e profissionais responsáveis pela concepção, planejamento, construção e manutenção dos sistemas, quanto para os próprios estudantes que interagem com a aplicação. Romero & Ventura (2007) afirmam que a descoberta do conhecimento pode ser orientada a diferentes atores de acordo com os interesses particulares de cada um:

- **Orientada para estudantes:** nesse contexto, técnicas de mineração de dados são utilizadas para recomendar atividades, recursos e tarefas que favoreçam e melhorem a

aprendizagem do aluno. Isso pode ser feito por meio de sugestões de *links*, baseadas nas experiências de aprendizagem de outros estudantes, por exemplo;

- **Orientada para educadores:** objetiva fornecer *feedback* aos instrutores; avaliar a estrutura do conteúdo do curso e a sua eficácia no processo de aprendizagem. Também é foco de interesse classificar grupos de estudantes baseado em suas necessidades; descobrir padrões de aprendizagem regulares e irregulares; descobrir os erros frequentes; descobrir atividades que são mais eficazes; organizar o conteúdo de forma mais eficiente para o progresso do aluno;
- **Orientada para gestores educacionais:** o objetivo é ter parâmetros sobre como melhorar a eficiência do sistema e adaptá-lo ao comportamento de seus usuários; ter medidas sobre como melhor organizar os recursos institucionais (humanos e materiais); melhorar a oferta de programas educacionais, etc.

Atualmente, a EDM vem se estabelecendo como uma forte e consolidada linha de pesquisa tanto nos EUA, quanto na Europa; contudo, no Brasil, essa área ainda é pouco explorada, apesar de crescer no país o potencial para a pesquisa, o desenvolvimento e a aplicação da mineração em ambientes educacionais (BAKER et al., 2011).

Há na literatura, diversos trabalhos sobre mineração de dados educacionais. Kampff et al. (2008) utilizaram técnicas de mineração para identificar perfis de alunos com risco de evasão ou reprovação em ambientes virtuais de aprendizagem (AVA). Ao verificar que um aluno possui tais características/comportamentos, o sistema alerta o professor que poderá tomar decisões pedagógicas necessárias para que o aluno fique mais motivado, se engaje, se sinta apoiado e seguro na sua aprendizagem.

Na pesquisa de Dias et al. (2008), os autores objetivaram analisar a utilização do ambiente de ensino-aprendizagem LabSQL a partir da aplicação de métodos provenientes da mineração de dados denominados Árvore de Decisão e Redes Bayesianas e descobrir informações relevantes sobre o perfil do aluno com relação à utilização dessa tecnologia.

Prata et al. (2009) aplicaram técnicas de mineração de dados em atividades colaborativas realizadas em ambientes de educação a distância. No trabalho, os autores estudaram a relação entre os atos de colaboração e a aprendizagem dos estudantes. Os resultados obtidos apontaram que os discentes que aprendem satisfatoriamente o conteúdo apresentado pelo professor, possuem maior probabilidade de mostrar comportamentos inapropriados específicos, como insultos, durante o andamento do curso. Tal resultado é interessante, pois muitos educadores

acreditam que os estudantes com baixo desempenho escolar são os maiores responsáveis por atrapalhar a aprendizagem dos demais alunos, contradizendo os dados obtidos na pesquisa de Prata et al. (2009).

Correia et al. (2011), utilizaram a técnica de clusterização para formar turmas similares de recuperação paralela. O estudo permitiu agrupar os estudantes pelos problemas de aprendizagem apresentados, inclusive, separando em um grupo específico aqueles com dificuldades de interpretação, podendo, assim, formar uma turma interdisciplinar. Os resultados obtidos podem ser utilizados para apoiar a tomada de decisão na formação de grupos de estudantes mais homogêneos.

O trabalho de Sheard (2010) concentra-se no uso de técnicas estatísticas e explicita como elas podem ser aplicadas na análise de *logs* de arquivos recolhidos de ambientes de aprendizagem baseados na Web. Já Mazza (2010), faz uso da técnica de visualização em sistemas educacionais Web, permitindo um rápido descobrimento de padrões e regularidades de aprendizagem para representar modelos de usuários.

Há ainda, trabalhos relacionados à construção de ferramentas de mineração de dados que sejam fáceis de serem utilizadas por educadores e não especialistas na área de mineração. Nesse contexto, destacam-se as pesquisas de Pedraza-Perez et al. (2011) e Merceron et al. (2005).

Pesquisas na área de EDM também proporcionaram modelos automatizados que podem ser utilizados durante a interação dos alunos com a interface do sistema, possibilitando verificar automaticamente quando ocorrem comportamentos inadequados. Assim, essa funcionalidade permite que sistemas educacionais apresentem comportamentos inteligentes. Baker et al. (2006), por exemplo, desenvolveram um personagem que reage de acordo com o comportamento apresentado pelo aluno no ambiente. Quando o estudante interage de maneira adequada, o personagem faz sinal de positivo. Já quando tenta trapacear, por exemplo, solicitando ajuda ao sistema diversas vezes na tentativa de obter a resposta de determinado exercício sem ao menos tentar resolvê-lo, o personagem altera seu comportamento e tenta diagnosticar o conhecimento do aluno através de uma sequência de perguntas adicionais. Segundo os autores, o comportamento do personagem desenvolvido auxiliou o professor a identificar os estudantes que estavam aprendendo "corretamente" e também incentivou os discentes a manterem um comportamento adequado no processo de aprendizagem.

4. ESTUDO DE CASO

Um estudo de caso, segundo Gil (2006), é caracterizado pelo estudo profundo e exaustivo de um ou de poucos objetos, que permita o seu conhecimento amplo e detalhado, tarefa praticamente impossível mediante os outros tipos de delineamentos considerados. Ele pode ser utilizado em pesquisas com diferentes propósitos, tais como a exploração de situações da vida real cujos limites não estão claramente definidos.

Visando a obtenção de um nível mais profundo de compreensão do conhecimento dos estudantes e dificuldades de aprendizagem em disciplinas de programação a fim de serem obtidos subsídios para superar esta realidade, um estudo de caso foi realizado na disciplina Programação Orientada a Objetos, ministrada no primeiro semestre de 2010 no curso de Licenciatura em Computação da Universidade de Pernambuco, onde havia 33 estudantes matriculados. A abordagem adotada na disciplina foi além da avaliação convencional, baseada no rendimento alcançado pelo aluno em poucas provas. Adotaram-se, assim, avaliações formativas, ocorridas ao longo do processo de aprendizagem. Dentro deste escopo, neste trabalho, foram considerados os dados obtidos a partir de 11 sessões de avaliação, a saber: 5 listas de exercícios, 4 miniprovas e 2 projetos. As atividades ocorreram durante todo o semestre e contemplaram diferentes conceitos. Cada sessão era composta por “n” enunciados de problemas, totalizando 890 exercícios a serem avaliados pelo professor titular da disciplina. Todas essas atividades antecederam as duas avaliações somativas, ocorridas ao fim de cada bimestre, e a análise e interpretação destes dados podem ser exploradas a fim de fornecer ao docente um melhor discernimento sobre o conhecimento e dificuldades dos estudantes na aprendizagem de programação. As avaliações foram formuladas de modo a obter uma medida cognitiva, que deveria retratar o real desempenho do estudante na resolução de cada problema e seria resultado do processo de correção do professor.

Após a análise dos enunciados dos problemas, foi possível classificá-los de acordo com os níveis da Taxonomia de Bloom. Pelo fato de as questões não terem sido formuladas tendo como referência uma taxonomia de objetivos educacionais, uma limitação deste trabalho está na classificação, já que o professor-autor dessas tarefas poderia ter um objetivo diferente do apresentado nos enunciados e ter equivocadamente utilizado um verbete de um nível cognitivo que não desejava avaliar naquele momento nos estudantes, levando, assim, à classificação, de uma questão em um nível superior ou inferior ao desejado pelo docente. O Quadro 3 apresenta a categorização das questões que compuseram as avaliações formativas propostas aos

estudantes tendo em vista a Taxonomia de Bloom e, o Quadro 4, exibe a sequência temporal com que essas atividades foram realizadas.

Quadro 3 – Classificação das questões das avaliações formativas de acordo com a Taxonomia de Bloom

Avaliação Formativa		Nível da Taxonomia de Bloom					
		Lembrar	Entender	Aplicar	Analisar	Avaliar	Criar
Lista de Exercícios 1	Questão 1			X			
	Questão 2			X			
	Questão 3			X			
	Questão 4					X	
Lista de Exercícios 2	Questão 1						X
	Questão 2						X
	Questão 3						X
	Questão 4						X
	Questão 5						X
Lista de Exercícios 3	Questão 1	X					
	Questão 2		X				
	Questão 3		X				
	Questão 4						X
	Questão 5						X
	Questão 6						X
Lista de Exercícios 4	Questão 1		X				
	Questão 2		X				
	Questão 3				X		
	Questão 4		X				
	Questão 5						X
Lista de Exercícios 5	Questão 1						X
Miniprova 1	Questão 1						X
Miniprova 2	Questão 1						X
Miniprova 3	Questão 1						X
Miniprova 4	Questão 1						X
Projeto 1							X
Projeto 2							X

Fonte: Elaborado pelo autor

Ao final de cada bimestre, após a realização de avaliações formativas, foram realizadas avaliações somativas. Os 33 estudantes matriculados na disciplina realizaram tais avaliações somativas, sendo que os valores possíveis estavam na faixa entre 0 e 10. Para efeitos de análise, os estudantes foram agrupados em dois grupos distintos. Os que alcançaram uma média igual ou superior a 7, que é a média requerida para aprovação, foram rotulados como aprovados, e os

estudantes restantes, que obtiveram uma média inferior a 7 ou um número de faltas superior a 25%¹, foram rotulados como reprovados.

A análise das médias das avaliações somativas revelou uma taxa de sucesso de apenas 24,24% dentre os estudantes da disciplina. Tal taxa de sucesso demonstra claramente a dificuldade encontrada por estes estudantes. A partir desta realidade, passou-se a tentar identificar se, dentre os dados disponíveis com o histórico das avaliações formativas, haveria evidências que pudessem identificar os fatores que favorecem e dificultam a aprendizagem dos estudantes na disciplina Programação Orientada a Objetos.

Quadro 4 – Sequência temporal das avaliações formativas

Sequência Temporal	Avaliação Formativa	Unidade
1	Lista de Exercícios 1	I
2	Miniprova 1	
3	Lista de Exercícios 2	
4	Miniprova 2	
5	Lista de Exercícios 3	
6	Miniprova 3	
7	Projeto 1	
8	Lista de Exercícios 4	II
9	Miniprova 4	
10	Lista de Exercícios 5	
11	Projeto 2	

Fonte: Elaborado pelo autor

Para aplicação das técnicas de mineração, após o conhecimento do domínio, partiu-se para a etapa de preparação dos dados. Esta envolveu a seleção, o pré-processamento e limpeza e a transformação dos dados de maneira mais apropriada para a mineração. O Quadro 5 apresenta o conjunto de atributos proposto e utilizado neste estudo de caso. Junto com os atributos são exibidos sua descrição, tipo e o domínio de valores que eles podem assumir.

Quadro 5 – Atributos propostos para o mapeamento do conhecimento do estudante

Atributo	Descrição	Tipo de Dado	Domínio
IdAluno	Código que identifica o aluno participante da sessão de avaliação	Numeric	[1, n]
IdSessao	Código que identifica o número da sessão de avaliação e a sequência temporal	Numeric	[1, n]

¹ Três estudantes foram reprovados por faltas na disciplina Programação Orientada a Objetos ofertada no curso de Licenciatura em Computação da Universidade de Pernambuco em 2010.1.

IdItem	Código que identifica uma questão numa avaliação formativa	Nominal	[Q ₀₁ , Q _n]
Conteúdo	Conteúdo tratado no item avaliado	Nominal	[C ₀₁ , C _n]
NívelCognitivo	Nível cognitivo do item avaliado de acordo com a Taxonomia de Bloom	Nominal	LEM – Lembrar ENT – Entender APL – Aplicar ANA – Analisar AVA – Avaliar CRI – Criar
ConceitoItem	Grau de acerto do aprendiz no item avaliado	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
ConceitoAvaliacao	Nível de desempenho do aprendiz numa determinada avaliação formativa	Nominal	A – 8 ≥ nota ≤ 10 B – 6 ≥ nota < 8 C – 4 ≥ nota < 6 D – 2 ≥ nota < 4 E – 0 ≥ nota < 2 ? – Ausência de informação
RotuloSomativa	Rótulo da avaliação somativa	Nominal	APR – Aprovado REP – Reprovado

Fonte: Elaborado pelo autor

A partir do conjunto de atributos apresentado no Quadro 5, algumas questões podem ser formuladas na tentativa de descobrir conhecimento através dos algoritmos de reconhecimento de padrões. Eis alguns exemplos:

- Existe relação entre os níveis cognitivos da Taxonomia de Bloom e o desempenho do aprendiz?
- Existe relação entre o desempenho do aprendiz e a sequência cognitiva com que um conteúdo é trabalhado: *Os estudantes apresentam melhor desempenho, em determinado conteúdo, quando o mesmo é trabalhado primeiramente no nível Aplicar e posteriormente no nível Entender?*
- Existe relação entre o desempenho do aprendiz nas avaliações formativas e o rendimento na avaliação somativa?
- Que agrupamentos homogêneos de estudantes podem ser identificados a partir de seus desempenhos numa disciplina específica?

5. EXTRAÇÃO DE PADRÕES E ANÁLISE DE RESULTADOS

Para aplicar as técnicas de mineração de dados, utilizou-se neste estudo de caso a ferramenta WEKA, desenvolvida pela Universidade de Waikato da Nova Zelândia (WITTEN & FRANK, 2005). A seguir, são descritas as três tarefas de mineração de dados utilizadas neste trabalho: classificação, associação e agrupamento. A escolha dos algoritmos que implementam as técnicas utilizadas nas tarefas mencionadas, baseia-se nos casos de sucesso apresentados por trabalhos correlatos destacados na Seção 1 e na subseção 3.3 que demonstraram que os algoritmos J4.8, *Apriori* e *K-means* obtiveram um bom resultado.

5.1. TAREFA DE CLASSIFICAÇÃO

Para esta tarefa foi necessário escolher um atributo alvo e os atributos de predição. Assim, escolheu como atributo alvo o conceito obtido pelo aprendiz na avaliação, ou seja, o nível de desempenho do estudante numa determinada avaliação formativa: A, B, C, D, E ou ?. A questão por trás desta escolha foi “O que determina o desempenho do aprendiz numa avaliação formativa?” (os níveis cognitivos que sustentam cada questão – *item* – da avaliação, a sequência temporal com que as avaliações formativas são realizadas, entre outros).

O algoritmo de árvore de decisão escolhido foi o J4.8. Nas Figuras 5 e 6, é possível visualizar, respectivamente, a saída do algoritmo selecionado e a árvore de decisão gerada com base nos 890 registros utilizados.

Figura 5 – Saída do algoritmo J4.8 com atributo alvo *ConceitoAvaliacao*

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    WekaExcel-weka.filters.unsupervised.attribute.Remove-R1
Instances:   890
Attributes:  7
              IdSessao
              IdItem
              Conteudo
              NivelCognitivo
              ConceitoItem
              ConceitoAvaliacao
              RotuloSomativa
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

```

```

ConceitoItem = ? : ? (395.0/82.0)
ConceitoItem = A
| IdSessao <= 4: A (156.0/2.0)
| IdSessao > 4
| | IdSessao <= 5: B (57.0/32.0)
| | IdSessao > 5: A (12.0)
ConceitoItem = B
| NivelCognitivo = ANA: B (0.0)
| NivelCognitivo = APL: B (0.0)
| NivelCognitivo = AVA: B (0.0)
| NivelCognitivo = CRI: B (58.0/6.0)
| NivelCognitivo = ENT: C (8.0)
| NivelCognitivo = LEM: B (0.0)
ConceitoItem = C: C (64.0/8.0)
ConceitoItem = D
| NivelCognitivo = ANA: E (3.0)
| NivelCognitivo = APL: D (0.0)
| NivelCognitivo = AVA: D (0.0)
| NivelCognitivo = CRI: D (16.0/2.0)
| NivelCognitivo = ENT: E (9.0)
| NivelCognitivo = LEM: D (0.0)
ConceitoItem = E
| IdSessao <= 6: E (106.0/8.0)
| IdSessao > 6: C (6.0/1.0)

Number of Leaves :      19

Size of the tree :      25

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      741          83.2584 %
Incorrectly Classified Instances    149          16.7416 %
Kappa statistic                    0.7779
Mean absolute error                 0.0873
Root mean squared error             0.2129
Relative absolute error             33.8676 %
Root relative squared error         59.3443 %
Coverage of cases (0.95 level)     97.191 %
Mean rel. region size (0.95 level) 46.5543 %
Total Number of Instances          890

=== Detailed Accuracy By Class ===

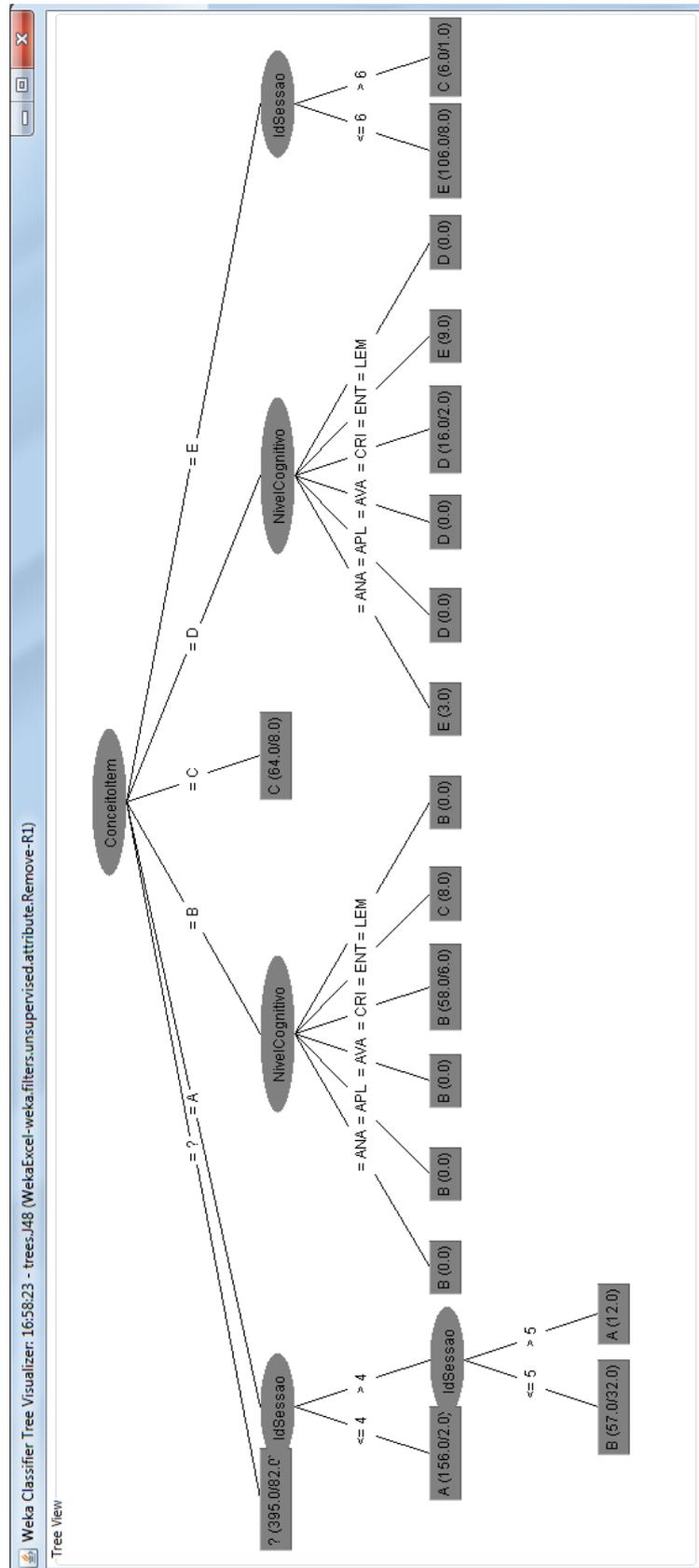
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
                0.997   0.142   0.792     0.997   0.883     0.822  0.916    0.754    ?
                0.843   0.009   0.965     0.843   0.9       0.877  0.961    0.929    A
                0.66    0.047   0.654     0.66    0.657     0.611  0.863    0.665    B
                0.657   0.015   0.848     0.657   0.74      0.719  0.93     0.721    C
                0.483   0.002   0.875     0.483   0.622     0.642  0.783    0.429    D
                0.78    0.013   0.917     0.78    0.843     0.82   0.921    0.79     E
Weighted Avg.   0.833   0.062   0.843     0.833   0.829     0.791  0.918    0.774

=== Confusion Matrix ===

  a  b  c  d  e  f  <-- classified as
313  0  0  1  0  0 |  a = ?
 12 167 19  0  0  0 |  b = A
 26  4  70  6  0  0 |  c = B
  9  0  16  67  0  10 |  d = C
  8  2  0  5  14  0 |  e = D
 27  0  2  0  2  110 |  f = E

```

Figura 6 – Árvore de decisão gerada com atributo alvo *ConceitoAvaliacao*



Fonte: Elaborada pelo autor

Analisando-se os padrões encontrados pelo algoritmo J4.8 representados na árvore da Figura 6, pode-se realizar as seguintes interpretações:

- Em 395 ocorrências, exceto em 82, quando o estudante obtém o conceito “?” (ausência de informação) no item, ele tem desempenho “?” na avaliação;
- Percebe-se que quando o estudante obtém conceito “A” (nota variando entre 8,0 e 10,0) no item, o desempenho final depende da sequência temporal da avaliação, sendo:
 - Obtém conceito “A” se o ID da sessão de avaliação for igual ou inferior a 4;
 - Se o ID da sessão de avaliação for superior a 4, então é necessário avaliar mais uma vez o ID da avaliação.
- Nota-se ainda que se o conceito obtido no item for igual a “B” (nota variando entre 6,0 e 8,0), então o conceito final na avaliação formativa dependerá do nível cognitivo em que o item foi classificado.
 - Em 58 casos, exceto em 6, obtém conceito “B” na avaliação formativa se o item foi classificado no nível Criar;
 - Em 8 casos, obtém conceito “C” na avaliação formativa se o item foi classificado no nível Entender.
- Em 64 ocorrências, exceto em 8, quando o estudante obtém o conceito “C” (nota variando entre 4,0 e 6,0) no item, ele tem desempenho final “C” na avaliação formativa.
- Observa-se, também, que se o estudante obtém conceito “D” (nota variando entre 2,0 e 4,0) no item avaliado, então o conceito final na avaliação formativa dependerá do nível cognitivo em que o item foi classificado:
 - Em 3 ocorrências, obtém conceito “E” na avaliação formativa se o item foi classificado no nível Analisar;
 - Em 16 casos, exceto em 2, obtém conceito “D” na avaliação formativa se o item foi classificado no nível Criar;
 - Em 9 casos, obtém conceito “E” na avaliação formativa se o item foi classificado no nível Entender.
- Verifica-se que se o estudante obtém conceito “E” (nota variando entre 0,0 e 2,0) no item avaliado, então o conceito na avaliação formativa dependerá da sequência temporal da avaliação, sendo:
 - Obtém conceito “E”, em 106 casos, exceto em 8, se o ID da sessão de avaliação for igual ou inferior a 6;

- Se o ID da sessão de avaliação for superior a 6, então o conceito na avaliação formativa é “C”.

Pode-se concluir com base nesta árvore de decisão que o desempenho pleno no item é determinante para o desempenho satisfatório do estudante na avaliação. Além disso, pode-se afirmar que o nível cognitivo em que o item é classificado, interfere no desempenho final do aprendiz na avaliação formativa. A sequência temporal das avaliações influencia também no desempenho final do estudante.

Após a descrição da árvore de decisão exibida na Figura 6, outro atributo foi selecionado para análise. Dessa vez escolheu-se como atributo alvo o rótulo na avaliação somativa (APR - Aprovado ou REP - Reprovado) e procurou-se responder a seguinte questão: Existe relação entre o desempenho do aprendiz nas avaliações formativas e o rendimento na avaliação somativa?

Nas Figuras 7 e 8, pode-se visualizar, respectivamente, a saída do algoritmo J4.8 e a nova árvore de decisão gerada.

Figura 7 – Saída do algoritmo J4.8 com atributo alvo *RotuloSomativa*

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    WekaExcel-weka.filters.unsupervised.attribute.Remove-R1
Instances:   890
Attributes:  7
              IdSessao
              IdItem
              Conteudo
              NivelCognitivo
              ConceitoItem
              ConceitoAvaliacao
              RotuloSomativa
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

ConceitoAvaliacao = ? : REP (314.0/50.0)
ConceitoAvaliacao = A
| IdSessao <= 5 : REP (178.0/53.0)
| IdSessao > 5 : APR (20.0/2.0)
ConceitoAvaliacao = B : REP (106.0/41.0)
ConceitoAvaliacao = C : REP (102.0/36.0)
ConceitoAvaliacao = D : REP (29.0/1.0)
ConceitoAvaliacao = E : REP (141.0/17.0)

Number of Leaves :    7

Size of the tree :    9

```

```

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      676           75.9551 %
Incorrectly Classified Instances    214           24.0449 %
Kappa statistic                     0.0186
Mean absolute error                 0.3553
Root mean squared error            0.4244
Relative absolute error             96.567 %
Root relative squared error        98.9861 %
Coverage of cases (0.95 level)     99.8876 %
Mean rel. region size (0.95 level) 99.2135 %
Total Number of Instances          890

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
                0.014   0.001   0.75       0.014   0.027     0.08  0.577    0.299    APR
                0.999   0.986   0.76       0.999   0.863     0.08  0.577    0.797    REP
Weighted Avg.   0.76   0.747   0.757     0.76   0.66      0.08  0.577    0.676

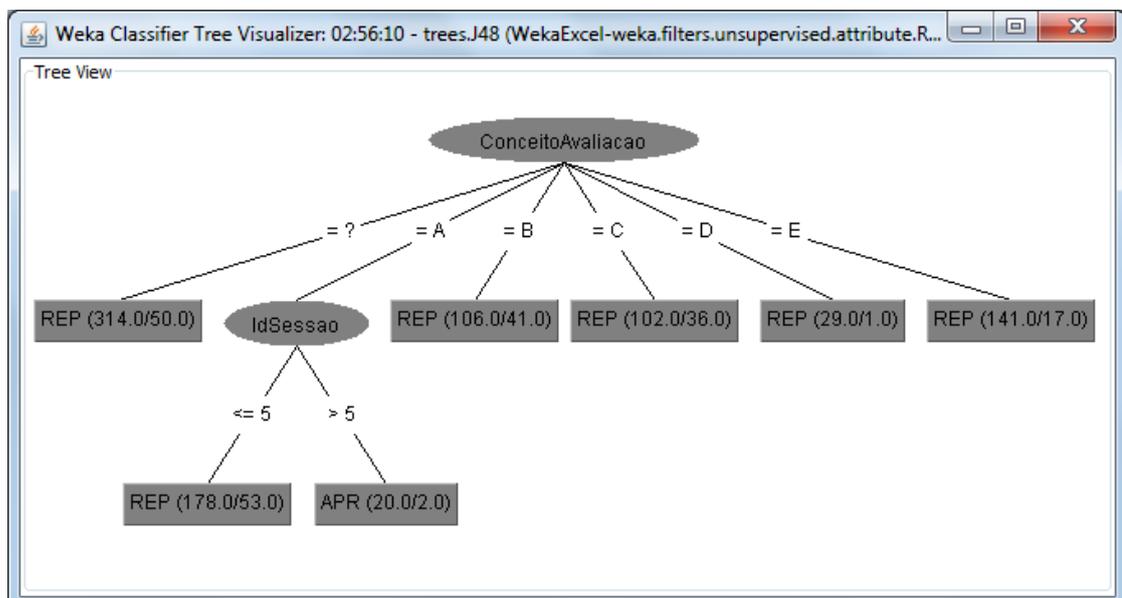
=== Confusion Matrix ===

 a  b  <-- classified as
 3 213 |  a = APR
 1 673 |  b = REP

```

Fonte: Elaborada pelo autor

Figura 8 – Árvore de decisão gerada com atributo alvo *RotuloSomativa*



Fonte: Elaborada pelo autor

Analisando-se os padrões encontrados pelo algoritmo J4.8 representados na árvore de decisão da Figura 8, pode-se realizar as seguintes leituras:

- Percebe-se que quando o estudante obtém o conceito “A” na avaliação formativa, o rótulo da avaliação somativa depende do ID da avaliação formativa, sendo:
 - Obtém rótulo APR (aprovado) se o ID da sessão de avaliação for superior a 5;
 - Obtém rótulo REP (reprovado) se o ID da sessão de avaliação for igual ou inferior a 5.
- Nota-se também que se o aprendiz obtém conceito “?”, “B”, “C”, “D” ou “E” na avaliação formativa, então o rótulo que ele recebe na avaliação somativa é REP (reprovado).

Pode-se concluir, com base nesses resultados, que, neste estudo, o desempenho pleno nas avaliações formativas 6, 7, 8, 9, 10 e 11² é determinante para a aprovação do estudante na disciplina Programação Orientada a Objetos.

Outros atributos alvos podem ser selecionados para fins de análises comparativas.

5.2. TAREFA DE ASSOCIAÇÃO

Na Figura 9 são apresentadas 9 regras geradas pelo algoritmo de regras de associação *Apriori*.

Figura 9 – Regras de associação geradas pelo algoritmo *Apriori*

```
Best rules found:
1. NivelCognitivo=CRI ConceitoAvaliacao=? 210 ==> ConceitoItem=? 210 <conf:(1)>
2. NivelCognitivo=CRI ConceitoAvaliacao=? RotuloSomativa=REP 168 ==> ConceitoItem=? 168 <conf:(1)>
3. Conteudo=Estruturas de Repeticao 132 ==> NivelCognitivo=CRI 132 <conf:(1)>
4. Conteudo=Estruturas de Repeticao RotuloSomativa=REP 100 ==> NivelCognitivo=CRI 100 <conf:(1)>
5. ConceitoAvaliacao=? 314 ==> ConceitoItem=? 313 <conf:(1)>
6. ConceitoAvaliacao=? RotuloSomativa=REP 264 ==> ConceitoItem=? 263 <conf:(1)>
7. ConceitoAvaliacao=A RotuloSomativa=REP 127 ==> ConceitoItem=A 119 <conf:(0.94)>
8. Conteudo=Arrays ConceitoItem=? 107 ==> NivelCognitivo=CRI 99 <conf:(0.93)>
9. ConceitoAvaliacao=A 198 ==> ConceitoItem=A 180 <conf:(0.91)>
```

Fonte: Elaborada pelo autor

Analisando-se os padrões as seguintes interpretações, dentre outras, podem ser feitas:

² Na disciplina Programação Orientada a Objetos ofertada no curso de Licenciatura em Computação da Universidade de Pernambuco em 2010.1, 11 avaliações formativas foram realizadas. Ver Quadro 4.

- Na regra 1, com confiança de 100% (conf =1), tem-se que quando o nível cognitivo é Criar e o conceito obtido pelo aprendiz na avaliação é “?” (ausência de informação), ele obtém esse mesmo conceito no item avaliado;
- Na regra 3, também com confiança de 100%, tem-se, 132 casos, que quando o conteúdo é Estruturas de Repetição, o nível cognitivo explorado é o Criar.
- Na regra 4, com 100% de confiança, percebe-se que quando o conteúdo é Estruturas de Repetição e o aprendiz obtém rótulo Reprovado na avaliação somativa, a abordagem de ensino fez uso do nível Criar.
- Na regra 8, com confiança de 93%, tem-se que quando o conteúdo é Arrays e não houve fornecimento de informação, pelo aprendiz, na resolução de um item que trate sobre esse conceito, o nível cognitivo empregado no questionamento foi o Criar.

5.3. TAREFA DE AGRUPAMENTO

A Figura 10 apresenta uma visão de 6 grupos obtidos com o algoritmo de agrupamento (clusterização) *K-means* na ferramenta WEKA, geradas com base nos 890 registros transformados a partir dos dados originais da disciplina Programação Orientada a Objetos.

Este algoritmo foi aplicado com a finalidade de permitir a identificação dos diferentes perfis de padrão de desempenho de estudantes, com base nos dados coletados sobre seu desempenho nas diversas avaliações formativas realizadas na disciplina.

Figura 10 – Agrupamento em *clusters*

Attribute	Full Data (890)	Cluster#		
		0 (113)	1 (144)	2 (351)
Conteudo	Arrays	Arrays	Arrays	Estruturas de Repeticao
NivelCognitivo	CRI	ENT	CRI	CRI
ConceitoAvaliacao	?	C	B	?
RotuloSomativa	REP	APR	APR	REP

	3 (122)	
=====		
Componentes Graficos do Pacote GUI + Objetos e Classes	APL	
	A	
	REP	
	4 (73)	5 (87)
=====		
Estruturacao de Sistemas em Camadas		Arrays
	ENT	CRI
	E	E
	REP	REP

Fonte: Elaborada pelo autor

Analisando os clusters da Figura 10 é possível realizar algumas interpretações:

- *Cluster 0*: Neste *cluster* com 113 ocorrências de itens (13%), foram agrupados os estudantes que alcançaram conceito “C” na avaliação formativa, composta por itens que tratavam sobre *Arrays*, abordados no nível Entender da Taxonomia de Bloom, e que obtiveram rótulo APR (aprovado) na disciplina;
- *Cluster 1*: Neste *cluster* com 144 ocorrências de itens (16%), foram agrupados os estudantes que alcançaram conceito “B” na avaliação formativa, composta por itens que abordavam o conteúdo *Arrays*, tratado no nível Criar da Taxonomia de Bloom, e que obtiveram rótulo APR (aprovado) na disciplina;
- *Cluster 2*: Este *cluster* agrupa os estudantes reprovados, que não responderam às questões (351 itens, perfazendo 39%) das avaliações formativas que tratavam sobre Estruturas de Repetição, as quais foram classificadas no nível Criar da Taxonomia de Bloom;
- *Cluster 3*: Este *cluster* agrupa os estudantes que realizaram satisfatoriamente as questões sobre Componentes Gráficos do Pacote GUI e Objetos e Classes (14% do total de itens), trabalhadas no nível Aplicar, mas que, no entanto, foram reprovados;
- *Cluster 4*: Foram agrupados, neste *cluster*, estudantes que obtiveram baixo desempenho na resolução de itens (8% de 890) que tratavam sobre Estruturação de

Sistemas em Camadas que foram trabalhados no nível Entender. No final da disciplina, os alunos deste grupo foram rotulados como reprovados.

- *Cluster 5*: Representa aqueles estudantes com baixo desempenho na resolução de itens (10% de 890) de avaliações formativas que tratavam sobre *Arrays*, abordados a partir do nível Criar. Os estudantes deste grupo foram rotulados como reprovados.

6. CONSIDERAÇÕES FINAIS

O processo contínuo de avaliação da aprendizagem pode gerar uma grande massa de dados que necessita ser tratada e interpretada de modo a fornecer informações precisas acerca do conhecimento dos aprendizes. Nesse sentido, técnicas de mineração de dados podem ser utilizadas para tal fim, como demonstrado neste trabalho.

Com a incorporação de referencial pedagógico junto às técnicas computacionais, aspectos específicos do processo de ensino-aprendizagem podem ser descobertos para tomada de decisões que favoreçam tal processo. Neste trabalho, buscou-se aporte teórico em Benjamin Bloom que, juntamente com outros pesquisadores, propuseram, em 1956, a Taxonomia de Bloom, uma taxonomia de objetivos educacionais que pode auxiliar os educadores em tarefas como planejamento e avaliação da aprendizagem.

Os resultados obtidos demonstram as potencialidades de técnicas de mineração de dados aliadas a objetivos educacionais. Em uma turma de Programação Orientada a Objetos, foi possível diagnosticar que existe relação entre os níveis cognitivos da Taxonomia de Bloom e o desempenho do aprendiz na disciplina, como também entre o desempenho do aprendiz nas avaliações formativas e o rendimento na avaliação somativa.

Através da análise de padrões de regras de associação, constatou-se que um dos gargalos na aprendizagem de Programação Orientada a Objetos está no processo de *criação*, a exemplo de estruturas de repetição e *arrays*. Este fator atrelado a outros, identificados em atividades de monitorias, como: i) a inabilidade do estudante em estabelecer uma nova rotina de estudos que lhe permita fazer uso de uma melhor gerência de tempo ou ii) a desmotivação do aprendiz em aprender os novos conceitos, muitas vezes tidos como de difícil compreensão, tem influenciado no aumento dos altos índices de reprovação nessa disciplina no curso de Licenciatura em Computação da Universidade de Pernambuco.

Através da técnica de clusterização conseguiu-se agrupar os estudantes pelos problemas de aprendizagem apresentados. Vale ressaltar, que ainda não é possível concluir se os resultados obtidos com essa técnica são significativos, entretanto, as informações descobertas poderão ser utilizadas para apoiar as reuniões pedagógicas e a tomada de decisão na formação de grupos de estudantes mais homogêneos.

Espera-se que, com os resultados dos padrões encontrados, estratégias pedagógicas sejam formuladas, a exemplo de atividades de monitoria dirigidas a grupos de aprendizes com tendência a não alcançar o desempenho requerido.

Como trabalhos futuros, pretende-se aprofundar os testes e análise dos padrões encontrados, bem como replicar o método utilizado em outras turmas e avaliar a intervenção dos resultados. Somado a isso, planeja-se incorporar ao conjunto de atributos proposto, outros dados que carreguem informações sobre o nível metacognitivo do aprendiz, como também sobre as estratégias por ele utilizadas na resolução das atividades propostas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, L. W.; KRATHWOHL, D. R. (Ed.). **A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives**. New York: Addison Wesley Longman, 2001.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação, v. 19, n. 02, p. 3-13, 2011.
- BAKER, R. S. J. D.; CORBETT, A. T.; KOEDINGER, K. R.; EVENSON, S. E.; ROLL, I.; WAGNER, A. Z.; NAIM, M.; RASPAT, J.; BAKER, D. J.; BECK, J. **Adapting to When Students Game an Intelligent Tutoring System**. In Proceedings of the International Conference on Intelligent Tutoring Systems, 2006, p. 392-401.
- BLOOM, B.S. (Ed.); ENGELHART, M. D.; FURST, E. J.; HILL, W. H.; KRATHWOHL, D. R.. **Taxonomy of educational objectives: The classification of educational goals**. Handbook 1: Cognitive domain. New York: David McKay, 1956.
- CARVALHO, J. V. . **Reconhecimento de Caracteres Manuscritos Utilizando Regras de Associação**. Dissertação de Mestrado. UFCG/COPIN, 2000.
- CERNY, R. Z.. **Uma reflexão sobre a avaliação formativa na educação a distância**. UFSC, 2001.
- CHEN, M., HAN, J., UY, P. S., **Data Mining: an Overview from Database Perspective**. In: Proceedings of the Advances in Knowledge Discovery in Databases, 1996.
- CORREIA, C. F; PIMENTEL, E. P. . **Mineração de Dados na Formação de Turmas para a Recuperação Paralela na Educação Básica**. In: XXII Simpósio Brasileiro de Informática na Educação, 2011, Aracaju. Anais do XXII SBIE - XVII WIE, 2011. p.172-175.
- DIAS, M. M. ; FILHO., L. A. DA S. ; LINO, A. D. P. ; FAVERO, E. L. ; RAMOS, E. M. L. S. **Aplicação de Técnicas de Mineração de Dados no Processo de Aprendizagem na Educação a Distância**. In: XIX Simpósio Brasileiro de Informática na Educação, 2008, Fortaleza - CE. Anais do XIX SBIE, 2008. p. 105-114.
- FAYYAD, U.; GRINSTEIN, G. G.; WIERSE, A. . **Information Visualization**. In: Data Mining And Knowledge Discovery. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 2002.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.. **From Data Mining to Knowledge Discovery**. American Association for Artificial Intelligence. 1996.
- FERRAZ, A. P. do C. M.; BELHOT, R. V.. **Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais**. Gest. Prod., São Carlos, v. 17, n. 2, p. 421-431, 2010.
- GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 5. Ed. 7. Reimpressão. São Paulo: Atlas, 2006.
- GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. . **Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais**. In: XXIII Simpósio Brasileiro de Informática na Educação, 2012, Rio de Janeiro. Anais do XXIII SBIE, 2012.
- HADJI, C.. **A avaliação desmistificada**. Porto Alegre: ArtMed, 2001.

- HAN, J.; KAMBER, M. . **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2001.
- HOFFMANN, J. M. L. **Avaliação: mito e desafio – uma perspectiva construtivista**. Porto Alegre: Mediação, 2010.
- JESUS, E. A. de; RAABE, A. L. A.. **Interpretações da Taxonomia de Bloom no Contexto da Programação Introdutória**. In: XX Simpósio Brasileiro de Informática na Educação, 2009, Florianópolis. Anais do XX SBIE, 2009.
- KAMPPFF, A. J. C.; REATEGUI, E. B.; LIMA, J. V. DE. **Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente**. Novas Tecnologias na Educação. V. 6, Nº 2, Dezembro, 2008.
- KRATHWOHL, D. R. **A revision of Bloom's taxonomy: an overview**. Theory in Practice, v. 41, n. 4, p. 212-218, 2002.
- LOPES, C. C. **Um Sistema de Apoio à Tomada de Decisão no Acompanhamento do Aprendizado em Educação a Distância**. Dissertação de Mestrado, UFCG. Campina Grande/PB, 2003.
- MAZZA, R. Visualization in Educational Environments. In: ROMERO, C., VENTURA, S., PECHENIZKY, M., BAKER, R.. **Handbook of Educational Data Mining**. 2010. Editorial Chapman and Hall/CRC Press, Taylor & Francis Group. Data Mining and Knowledge Discovery Series.
- MERCERON, A.; YACEF, K. **TADA-Ed for Educational Data Mining**. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning. Volume 7, Number 1, May, 2005.
- PEDRAZA-PEREZ, R.; ROMERO, C.; VENTURA, S. **A Java desktop tool for mining Moodle data**. In Proceedings of the 4th International Conference on Educational Data, 2011.
- PERRENOUD, P.. **Avaliação: da excelência à regulação das aprendizagens entre duas lógicas**. Porto Alegre: Artes Médicas, 1999.
- PIMENTEL, E. P.; OMAR, N.. **Descobrendo Conhecimentos em Dados de Avaliação da Aprendizagem com Técnicas de Mineração de Dados**. In: XII Workshop de Informática na Escola, 2006, Campo Grande, MS. Anais do XXVI Congresso da Sociedade Brasileira de Computação, 2006.
- PRATA, D. N.; BAKER, R. S. J. D.; COSTA, E., ROSÉ, C. P.; CUI, Y.; CARVALHO, A. M. J. B de. **Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments**. In Proceedings of the International Conference on Educational Data Mining, 2009. p. 131-140.
- QUINLAN, R. . **C4.5: Programs for Machine Learning**, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- RAPKIEWICZ, C. E; FALKEMBACH, G; SEIXAS, L.; ROSA, N. S.; CUNHA, V. V. da; KLEMMANN, M. **Estratégias pedagógicas no ensino de algoritmos e programação associadas ao uso de jogos educacionais**. Revista Novas Tecnologias na Educação, v. 4, nº 2, Dezembro, 2006.
- REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; PAULA, M. F. de. Mineração de dados. In: REZENDE, S. O. . **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri, SP: Manole, 2003.

ROMERO, C; VENTURA, S.. **Educational Data Mining: A Survey from 1995 to 2005**. Expert Systems with Applications 33, 2007, p. 125-146.

SANTOS, H. L. dos; CAMARGO, F. N. P.; CAMARGO, S. da S.. **Minerando Dados de Ambientes Virtuais de Aprendizagem para Predição de Desempenho de Estudantes**. In: Seventh Latin American Conference on Learning Objects and Technologies, 2012, Guayaquil, Equador. Proceedings of the 7th LACLO, 2012.

SCOTT, T.. **Bloom's taxonomy applied to testing in computer science classes**. In: Journal of Computing Sciences in College, USA, 2003, v. 19, n. 1, p. 267-274.

SHEARD, J. Basics of Statistical Analysis of Interactions Data from Web-Based Learning Environments. In: ROMERO, C., VENTURA, S., PECHENIZKY, M., BAKER, R. **Handbook of Educational Data Mining**. 2010. Editorial Chapman and Hall/CRC Press, Taylor & Francis Group. Data Mining and Knowledge Discovery Series.

THOMPSON, E.; LUXTON-REILLY, A.; WHALLEY, J. L.; HU, M.; ROBBINS, P.. **Bloom's taxonomy for CS assessment**. In: X Australasian Conference on Computing Education. Proceedings of the ACE'08, 2008, v. 78, p. 155-161.

UNIVERSITY OF WAIKATO. **Weka 3 - Data Mining with Open Source Machine Learning Software in Java**. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em 14 de set. 2012.

WHALLEY, J. L.; LISTER, R.; THOMPSON, E.; CLEAR, T.; ROBBINS, P.; KUMAR, P. K. A.; PRASAD, C.. **An Australasian Study of Reading and Comprehension Skills in Novice Programmers, using the Bloom and SOLO Taxonomies**”, In: 8th Australasian Conference on Computing Education. Proceedings of the ACE '06, 2006, v. 52, p. 243-252.

WITTEN, I. H.; FRANK, E.; HALL, M. A.. **Data Mining: Practical Machine Learning Tools and Techniques**, 3rd edition, Morgan Kaufmann, Burlington, MA, 2011,

WITTEN, I. H.; FRANK, E. . **Data Mining: Practical Machine Learning Tools and Techniques**. 2nd edition, Morgan Kaufmann Publishers, San Francisco, CA, 2005.