



A contribuição de Mineração de Dados no processo de Autoavaliação dos cursos superiores do Instituto Federal de Sergipe

Leopoldo Ramos de Oliveira¹, Fausto Bernard Melo Soares², José Acácio de Jesus Moura³, Patrick Henrique da Silva Brito⁴, Aydano Pamponet Machado⁵

¹Mestrando em Modelagem Computacional de Conhecimento – UFAL. e-mail: leopoldoroliveira@gmail.com

²Mestrando em Modelagem Computacional de Conhecimento – UFAL. e-mail: fausto.ifs@gmail.com

³Mestrando em Modelagem Computacional de Conhecimento – UFAL. e-mail: acaciomoura@ibest.com.br

⁴Doutor em Ciência da Computação – UNICAMP. e-mail: patrick@ic.ufal.br

⁵Doutor em Informatique – Laboratoire D'informatique de Paris VI - e-mail: aydano.machado@ic.ufal.br

Resumo: Este artigo tem como foco avaliação institucional que é um dos componentes do instrumento da avaliação das instituições de ensino superior (IES), previsto no SINAES (Sistema Nacional de Avaliação da Educação Superior), bem como a análise dos resultados da avaliação realizada pelos discentes quanto aos cursos superiores do IFS (Instituto Federal de Sergipe) no processo de autoavaliação desta instituição no contexto da IA (inteligência artificial), utilizando a mineração de dados e a técnica de classificação através do algoritmo, árvore de decisão (decision tree) implementado no software Rapidminer (versão 5.2). Através da Mineração de dados foi possível descobrir a relevância dos atributos (questões) envolvidos no processo de autoavaliação com o objetivo de detectar fragilidades e potencialidades da IES e fornecer aos gestores subsídios para tomada de decisão

Palavras-chave: Avaliação, SINAES, Mineração de dados, Árvore de Decisão

1. INTRODUÇÃO

O SINAES (sistema nacional de avaliação da educação superior) foi instituído pela lei 10.861, de 14 de abril de 2004 com os objetivos de promover a melhoria da qualidade da educação superior e orientar a sua expansão bem como oferecer as instituições de ensino superior (IES) a oportunidade do autoconhecimento acerca das suas potencialidades e fragilidades. O processo de autoconhecimento é obtido através da avaliação interna realizada pelas IES de acordo com as normas estabelecidas no roteiro de autoavaliação que foi concebido a partir da lei supra mencionada.

As IES encontram dificuldades para análise e interpretação dos dados obtidos durante o processo de autoavaliação, pois o mesmo contém uma grande base de dados tendo em vista que toda a comunidade acadêmica participa da autoavaliação. Nesse contexto, o processo de descoberta de conhecimento (KDD) pode contribuir com a descoberta de informações muitas vezes “escondidas” nessas bases de dados, pois quando se utilizam apenas as técnicas triviais da estatística as mesmas normalmente não são percebidas. A mineração de dados como a principal etapa do KDD utiliza técnicas automáticas para extrair novos padrões e relações entre as variáveis (atributos) da base analisada. Em geral, as técnicas utilizadas são: Análise de agrupamentos, Estatísticas descritiva, Regras associativas e Árvore de decisão.

O presente trabalho utiliza o método de pesquisa experimental através do software RAPID MINER versão 5.2, um ambiente para aprendizagem de máquina e realização de mineração de dados com o objetivo de verificar as vantagens e desvantagens de modelos criados a partir dos algoritmos de classificação. O processo de descoberta de conhecimento gerado através dos atributos da base obedece as seguintes fases: Seleção dos dados, Limpeza dos dados, Enriquecimento dos dados, Transformação dos dados, Mineração dos dados e por fim a Apresentação e Análise dos dados.

ISBN 978-85-62830-10-5

VII CONNEPI©2012



A mineração de dados deste trabalho tem como foco a avaliação dos cursos superiores com o objetivo de descobrir novos conhecimentos e dessa forma contribuir para a tomada de decisão da IES visando corrigir fragilidades e manter ou melhorar as potencialidades rumo ao ensino de excelência.

1.1 - AVALIAÇÃO INTERNA OU AUTOAVALIAÇÃO.

A avaliação interna ou autoavaliação é um processo que permite que a IES construa o conhecimento sobre sua realidade com o objetivo de cumprir com qualidade o seu papel no contexto social onde está inserida. A análise e interpretação dos dados obtidos na avaliação visam detectar potencialidades e fragilidades da IES e servir como referencial para a tomada de decisão no sentido de corrigir os pontos fracos da instituição e implementar políticas educacionais que atendam as demandas da comunidade. Esses dados também são utilizados para retroalimentar o planejamento estratégico da IES. (SINAES, p. 11, 2004)

“A avaliação é um processo cíclico, criativo, e renovador de análise, interpretação e síntese das dimensões que definem a IES” (SINAES, 2004, p. 11)

1.1.1 - REQUISITOS DA AUTOAVALIAÇÃO

Para que a autoavaliação alcance bons resultados é imprescindível que as seguintes condições sejam atendidas em sua totalidade. (SINAES, p. 12, 2004)

- Existência de uma equipe de coordenação;
- Participação de integrantes de todos os segmentos da IES;
- Compromisso explícito dos gestores da IES;
- Informações, coleta e dados sejam confiáveis;
- Uso efetivo dos resultados da avaliação

1.1.1.1 - ETAPAS DA AVALIAÇÃO INTERNA

Os resultados da autoavaliação serão submetidos aos especialistas no momento da avaliação externa designada pelo MEC. Assim, é preciso que esses dados reflitam a realidade da IES e que todas as ações decorridas face as avaliações internas estejam bem evidenciadas no relatório de autoavaliação e sejam amplamente divulgadas para a comunidade. Recomendações de especialistas no momento da visita externa podem ajudar a IES a consolidar o seu processo de autoavaliação (SINAES, 2004)

Para garantir um processo de avaliação interna que atenda as exigências do MEC é necessário que ocorra as seguintes etapas:

- 1º etapa: **Preparação**
 - constituição de uma comissão composta por membros da comunidade acadêmica;
 - sensibilização de todos os participantes da autoavaliação ;
 - Elaboração do projeto de avaliação;



- 2ª Etapa: **Desenvolvimento**

- Coerência entre as ações planejadas, a metodologia adotada e articulação entre os participantes;
- Levantamento de dados e informações;
- Análise das informações e elaboração de relatórios parciais;

- 3ª Etapa: **Consolidação**

- Relatório final contendo uma análise crítica e isenta apontando ações de melhorias para IES;
- Divulgação dos resultados da avaliação a comunidade acadêmica ;
- Balanço crítico do processo de auto avaliação.

1.2 – O PROCESSO DE DESCOBERTA DE CONHECIMENTO (KDD)

Criado para auxiliar o ser humano em descobrir informações ocultas em bases de dados, tais como tipos de padrões comportamentais dos dados, o processo de “descoberta de conhecimento em bases de dados” (KDD – Knowledge Discovery in Database) (Fayyad, 1996) transforma dados em informação, podendo ser analisados e desta forma se tornar conhecimento útil. Segundo Carvalho (2001), esse processo é interdisciplinar e envolve diversas áreas, entre elas, estatística e matemática, banco de dados, aprendizado de máquina e reconhecimento de padrões.

O processo de KDD, vislumbrado na Figura 1, é composto das atividades de Seleção, Processamento, Transformação, Mineração de dados e Interpretação (Kumar, 2006) e são definidas como:

1. Seleção de dados: trata-se da seleção e coleta do conjunto de dados necessários à descoberta de conhecimento.

2. Pré-Processamento e limpeza (Processamento): aqui são removidas as informações desnecessárias e são definidas estratégias para o tratamento de informações não consistentes e irrelevantes.

3. Transformação: nesta etapa objetiva-se contornar eventuais limitações presentes em algoritmos de descoberta de padrões que serão utilizados na fase de mineração de dados. Para isso os dados são transformados para um formato adequado para aplicação do algoritmo selecionado. Dentre os tipos mais comuns pode-se citar a normalização e discretização de atributos quantitativos.

4. Mineração de dados: é a etapa do processo de descoberta de conhecimento foco do presente trabalho, na qual ocorre a extração de padrões que não estão explícitos nos dados trabalhados. Esses padrões podem ser extraídos através de técnicas de classificação, agrupamento ou sumarização (Fayyad, 1996).

5. Interpretação: neste momento os resultados obtidos são avaliados e interpretados cujo objetivo é a transformação em conhecimento útil para o apoio a decisões.

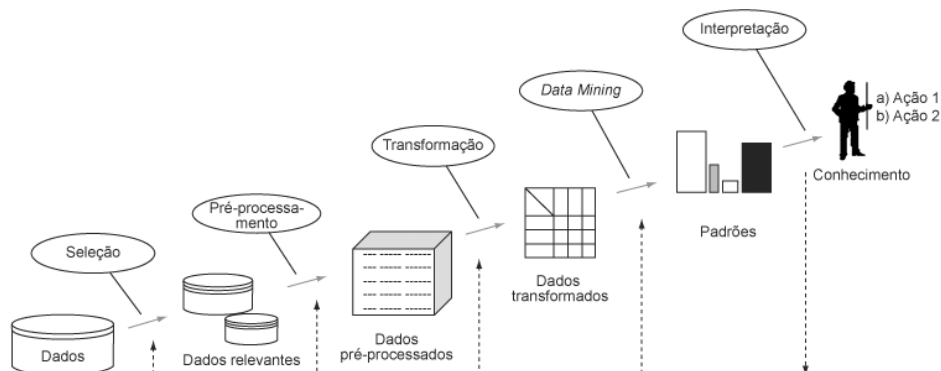


Figura 1 - Etapas do processo de KDD. Fonte: FAYYAD, 1996

Algumas tarefas podem ser definidas como foco da mineração de dados, são elas: classificação, associação e agrupamento (ou clusterização) (KUMAR et al. , 2006). A classificação trata-se da alocação de dados a partir de um conjunto de dados previamente utilizados como entrada ao classificador chamado de base de treinamento. Assim, a tarefa de classificação é dividida em duas etapas: fase de treinamento e fase de teste. Na primeira, os dados são apresentados ao algoritmo de aprendizagem para que este, por sua vez, efetue a construção um modelo de classificação. Enquanto que na segunda etapa, o modelo é submetido a dados ainda desconhecidos pelo mesmo, e tem a finalidade de alocar esses dados em classes (HAN, KAMBER,2006).

Para medição de desempenho do algoritmo de aprendizagem utilizado na mineração de dados é a matriz da confusão, que trata-se de uma ferramenta de visualização usada tipicamente dentro aprendizagem supervisionada (em não supervisionada aprendizagem é chamado tipicamente a matriz combinando). Cada coluna da matriz representa os exemplos em uma classe predita pelo classificador, quando cada linha representar os exemplos em uma classe real. Um benefício de uma matriz da confusão é que é fácil ver se o sistema estiver confundindo duas classes (isto é. geralmente mislabelling um como outro).

2. MATERIAL E MÉTODOS

No presente trabalho foi utilizado o algoritmo C4.5 (Quinlan, 1993) para construção de árvore de decisão, que é uma representação simples utilizada por vários sistemas de aprendizagem de máquina, e assim efetuar a classificação de dados, objetivando a busca pela relevância dos atributos envolvidos no processo de autoavaliação. O método prático utilizado será a pesquisa experimental, na qual foram feitos experimentos baseados nas aplicações do algoritmo.

2.1- Etapas do processo de descoberta de conhecimento na base de dados da Avaliação Institucional do IFS do ano de 2012

Durante o processo de descoberta de conhecimento seguiu-se as 4 etapas descritas no item 1.2, que podem ser vislumbradas detalhadamente a seguir.

2.1.1 Etapa 1: Seleção de dados

Nessa etapa foi obtida a base de dados da Avaliação Institucional do Instituto Federal de Sergipe do ano de 2012, estando a mesma com a extensão .sav, fruto do software SPSS, até então



utilizado na instituição para realizar a tarefa de tabulação dos dados bem como a análise estatística dos mesmos.

Assim a base de dados escolhida para participação neste processo se refere a avaliação dos discentes quanto ao seu curso, cujos instrumentos de avaliação possuem como repostas notas variando de 1 a 5, onde:

- 1: Desconheço;
- 2: Ruim;
- 3: Regular;
- 4: Bom;
- 5: Muito Bom;

2.1.2 Etapa 2: Limpeza e Transformação dos dados

De posse da escolha do domínio a ser abordado (Avaliação Institucional, mais precisamente avaliação do discente de curso superior no Instituto Federal de Sergipe) e obtida a base de dados, foi executada a limpeza dos dados, para a definição de campos (atributos) relevantes ao processo de Descoberta de Conhecimento e conversão do rótulo dos mesmos, da modalidade de questões numeradas para atributos nominais, ficando assim mais significativos.

Também foi criado um atributo chamado Avaliação Curso para atuar como classe na base de conhecimento, atributo cujo qual será classificado o Curso em: Ruim, Regular, Bom, Muito Bom.

A base resultante desse processo contém 321 registros rotulados de acordo com os seguintes atributos:

- Estrutura Curricular;
- Carga Horária das Disciplinas;
- Carga Horária Total;
- Laboratórios Informática;
- Número Docentes;
- Qualificação Docentes;
- Promoção de Extensão;
- Envolvimento Prof./Pesquisa;
- Envolvimento Alunos/Pesquisa;
- Articulação de Conteúdos;
- Laboratórios Específicos;
- Aulas Práticas de campo;
- Curso e Mercado de Trabalho;
- Contribuição Disciplinas para Formação;
- Avaliação Curso (Classe);

2.1.3 Etapa 3: Mineração

Através do software Rapid Miner 5.2, um ambiente de trabalho com algoritmos de aprendizagem de máquina, foi realizada a mineração de dados, etapa principal do processo de Descoberta de Conhecimento. A ferramenta supracitada fornece recursos de importação de dados, algoritmos de classificação de dados, algoritmos de agrupamento, entre outros de grande utilidade no processo de descoberta de conhecimento.

No presente trabalho foram utilizados os operadores: Decision Tree, Apply Model, Performance, Weight by Relief.

Assim, construiu-se um modelo (Figura 2) de árvore de decisão aplicando o algoritmo C4.5 e posteriormente verificou-se a sua exatidão quanto a predição da classe (Avaliação Curso) da base de conhecimento, através de um recurso chamado matriz de confusão.

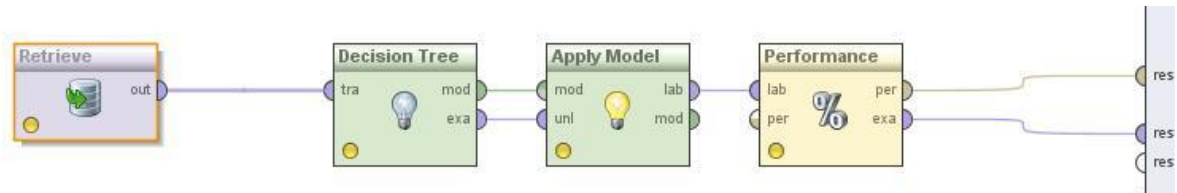


Figura 2: Processo no rapid Miner para construção da Árvore de Decisão

Em seguida, na Figura 3, foi calculado o peso relativo aos atributos (questões), e gerada a tabela 2 com os valores já normalizados.

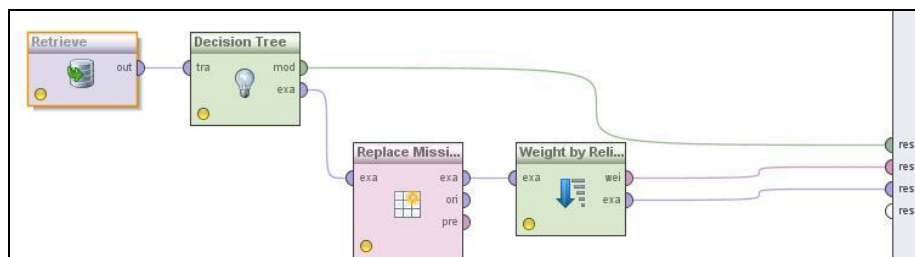


Figura 3 - Calculo dos pesos dos atributos no Rapid Miner.

3. RESULTADOS E DISCUSSÕES

A matriz de confusão da tabela 1 revela que a árvore de decisão representa adequadamente os dados obtidos na avaliação institucional ,tendo em vista a precisão de 94,39% do classificador.Ressalta-se que o conceito muito bom teve uma previsão de 71,43% devido a baixa na base em estudo. A tabela 2 mostra os pesos relativos aos atributos da base revelando a importância dos mesmos no processo de autoavaliação.

Tabela 1 - Matriz de Confusão

Accuracy: 94,39%					
	True Bom	True Regular	True Muito Bom	True Ruim	Class precision
Pred. Bom	153	3	2	0	96,84%
Pred. Regular	6	131	0	1	94,93%
Pred. Muito Bom	1	0	5	0	83,33%
Pred. Ruim	2	3	0	14	73,58%
Class Recall	94,44%	95,62%	71,43%	93,33%	



Tabela 2 - Matriz com pesos e normalização dos Atributos.

Atributo	Peso	Normalização	Frequência
Qualificação Docentes	0	0	0,00%
Carga Horária Total	0,34	0,05	5,23%
Carga Horária Disciplinas	0,06	0,01	0,92%
Promoção de Extensão	0,249	0,04	3,83%
Laboratórios Específicos	0,247	0,04	3,80%
Estrutura Curricular	0,431	0,07	6,63%
Contribuição Disciplinas para Formação	0,436	0,07	6,71%
Aulas Práticas de Campo	0,483	0,07	7,43%
Número Docentes	0,517	0,08	7,96%
Curso e Mercado de Trabalho	0,565	0,09	8,69%
Articulação de Conteúdos	0,635	0,10	9,77%
Laboratórios de Informática	0,711	0,11	10,94%
Envolvimento Prof/Pesquisa	0,825	0,13	12,69%
Envolvimento Alunos/Pesquisa	1	0,15	15,39%

6. CONCLUSÕES

Após análise dos resultados obtidos através da mineração de dados com a utilização do Rapid Miner foi necessário normalizar os dados da tabela.. para um melhor entendimento dos resultados obtidos visando dar semântica aos mesmos e identificar os problemas de relevância da avaliação dos discentes quanto aos cursos do IFS. Os pesos atribuídos aos atributos da base deve servir como estudo por parte dos gestores do IFS como ferramenta de apoio a decisão.

REFERÊNCIAS

BRASIL. Decreto-Lei n.10.861, de 14 de abril de 2004. **Institui o Sistema Nacional de Educação Superior – SINAES** - e dá outras providências.

_____. Portaria n. 2.501, de 9 de julho de 2004. **Regulamenta os procedimentos de avaliação do Sistema nacional de Avaliação da Educação Superior (SINAES)**, instituída na Lei n. 10.861 de 14 de abril de 2004.

BRASIL. MEC. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Comissão Nacional de Avaliação da Educação Superior. **Roteiro de auto-avaliação institucional**. Brasília: Editora, 2004.

FAYYAD, et al. **Advances in Knowledge and data mining**. AAAI Press/MIT Press, 1996.



HAN, Jiawei; KAMBER, Micheline. **Data Mining Concepts and Techniques**. 2. Ed. San Francisco: Elsevier, 2006

KUMAR, Vipin; STEINBACH, Michael; TAN, Pang-Ning; **Introduction to Data Mining**. Pearson Education, 2006.

MITCHELL, TOM. **Machine Learning**. New York: McGraw-Hill, 1997.