

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
SISTEMAS DE INFORMAÇÃO**

**MINERAÇÃO DE DADOS EM DADOS DO
CURSO PRÓ-CONSELHO/UFSM PARA A
IDENTIFICAÇÃO DO PERFIL DOS
CONSELHOS MUNICIPAIS DE EDUCAÇÃO
DO RS**

PROJETO DE TRABALHO DE GRADUAÇÃO

Gabriel Machado Lunardi

Santa Maria, RS, Brasil

2014

**MINERAÇÃO DE DADOS EM DADOS DO CURSO
PRÓ-CONSELHO/UFSM PARA A IDENTIFICAÇÃO DO
PERFIL DOS CONSELHOS MUNICIPAIS DE EDUCAÇÃO DO
RS**

Gabriel Machado Lunardi

Projeto de Trabalho de Graduação a ser apresentado ao curso de Sistemas de Informação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de
Bacharel em Sistemas de Informação

Orientador: Prof.^a Dr.^a Ana T. Winck

Santa Maria, RS, Brasil

2014

1 INTRODUÇÃO

O Programa do Pró-Conselho é proposto pelo Ministério da Educação (MEC), consiste em um Curso de Extensão a Distância de Formação Continuada de Conselheiros Municipais de Educação. Segundo (LUNARDI; LUNARDI, 2014), o curso configura-se como iniciativa da Secretaria de Educação Básica (SEB) que visa fortalecer os Sistemas de Ensino e as instâncias políticas e sociais tal como é o Conselho Municipal de Educação (CME). Tem carga horária de 160h, ofertado via internet, em ambiente virtual de aprendizagem (plataforma *Moodle*) ministrado pela Universidade Federal de Santa Maria em parceria com a Coordenação do Programa Nacional de Capacitação de Conselheiros Municipais de Educação – Pró-Conselho, SEB/MEC. O público alvo é formado por Conselheiros Municipais de Educação e técnicos das secretarias de educação dos municípios onde ainda não existam Conselhos Municipais de Educação.

O Pró-Conselho/UFSM é responsável pela oferta do curso no estado do Rio Grande do Sul - RS, lotado no Centro de Educação da UFSM, estando em sua 2ª edição. A equipe é composta por: coordenação geral; coordenação adjunta local; coordenação adjunta pedagógica; um professor supervisor; tutores; apoio de Informática e apoio administrativo. Vale ressaltar que o projeto está diretamente ligado aos grupos de pesquisa dessa área, pois serve como objeto de estudo.

Nessa perspectiva, como uma das atividades do curso, a equipe gestora formulou um questionário composto por perguntas abertas e fechadas buscando delinear o perfil dos CME no RS. Tal questionário foi utilizado como requisito inicial no desenvolvimento de um banco de dados relacional o qual receberá as respostas. A partir daí, um sistema *web*, a ser desenvolvido, funcionará como instrumento de coleta e, após análise, visualização dos dados. É importante destacar que existe um sistema, desde 2003, com proposição parecida, porém de âmbito nacional, chamado Sistema de Informações dos Conselhos Municipais de Educação (SICME, 2014). Todavia, conforme consta no site (MEC, 2014) o sistema encontra-se em manutenção.

Sendo assim, como tema central, empregar-se-á técnicas de mineração de dados buscando padrões que auxiliem a identificar o perfil dos CME no RS através dos dados coletados via sistema *on-line*. Ao identificar o perfil, entende-se ser possível explorar a realidade dos CME e, por conseguinte, aplicar políticas e ações que melhorem tais órgãos fortalecendo o sistema de educação básica.

2 OBJETIVOS

2.1 Objetivo Geral

Empregar técnicas de mineração de dados, em um banco de dados relacional do curso Pró-Conselho/UFSM, buscando padrões que auxiliem na identificação do perfil e realidade dos CME no RS.

2.2 Objetivos Específicos

- Construir um banco de dados relacional e populá-lo;
- Desenvolver um sistema *web* para a coleta e visualização dos dados.
- Pesquisar sobre as técnicas de mineração de dados;
- Estudar formas de pré-processar o *data set* obtido;
- Identificar e aplicar uma técnica de mineração de dados que melhor se enquadre sobre os dados resultantes do pré-processamento;
- Encontrar padrões que auxiliem na identificação do perfil dos conselhos municipais de educação no RS.

3 JUSTIFICATIVA

As atividades cotidianas de qualquer pessoa implicam na geração de dados, os quais são armazenados por muitas instituições sumarizando grandes quantidades. Entretanto, muitas dessas instituições não atingem a transformação dos dados em conhecimento permitindo melhorar a qualidade de vida das pessoas.

Nesse sentido, com a identificação do perfil dos conselhos municipais de educação no RS, através da mineração de dados, será possível gerar conhecimento sobre a realidade dos CME. A partir disso, alcançar-se-á o refinamento do Pró-Conselho/UFSM e, conseqüentemente da instrumentalização dos cursistas que, por sua vez, estarão melhor preparados para implementar os conselhos e, principalmente, contribuir para a qualidade do ensino básico brasileiro.

4 REVISÃO DE LITERATURA

4.1 Mineração de Dados

Na atualidade são gerados muitos dados, principalmente em razão da rápida evolução computacional, onde destaca-se a disseminação da internet. Conforme mostra (REZENDE, 2005), esse grande volume de dados é armazenado em repositórios de diferentes áreas, seja médica, científica, financeira, comercial, dentre muitas outras que nos circundam. A partir disso, informações úteis podem ser extraídas dos dados, porém torna-se inviável extraí-las e interpretá-las manualmente, utilizando só a subjetividade humana. É nesse sentido, em automatizar parte do processo, que existem técnicas computacionais adequadas, em especial as relacionadas a mineração de dados.

Tendo em mente o exposto, a mineração de dados, em consonância com (HAND; MANNILA; SMITH, 2001), é "a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis enquanto compreensíveis ao dono dos dados".

Já na perspectiva de (CABENA et al., 1998), a mineração de dados é "um campo multidisciplinar que reúne técnicas de aprendizagem de máquina, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados".

Neste trabalho entende-se mineração de dados como uma etapa de um processo maior denominado KDD (*Knowledge Discovery in Databases* - Descoberta de Conhecimento em Bases de Dados), pois existem etapas, anteriores e posteriores a mineração, muito importantes. Alguns defensores desse ponto de vista são (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996) que reafirmam "[...] KDD refere-se a todo o processo de descoberta de conhecimento e que mineração de dados compreende a aplicação de algoritmos específicos para a extração de padrões". No entanto, na literatura, ainda não é consenso da definição de mineração de dados e KDD, porém há a consonância de que o procedimento de mineração deva ser iterativo, interativo e dividido em fases.

Vale mencionar que mineração de dados não compreende uma simples consulta em um banco de dados ou, então, uma pesquisa na *web*. Essas ações pertencem a área de recuperação de informação (*information retrieval*) e não devem ser confundidas com mineração de dados,

como verificado por (TAN; STEINBACH; KUMAR, 2006).

4.2 Tarefas da Mineração de Dados

As tarefas de mineração de dados são divididas em duas categorias: preditivas e descritivas. As preditivas preveem o valor de um atributo com base nos valores de outros atributos, são elas: classificação e estimação. Já as tarefas descritivas visam a derivação de padrões (correlações, tendências, grupos) que sintetizem os relacionamentos subjacentes nos dados, sendo: regras de associação, agrupamento, dentre outras (TAN; STEINBACH; KUMAR, 2006). Assim, (LAROSE, 2005) elenca as principais tarefas conforme explicado.

Classificação (Classification): identifica a qual classe (categoria) pertence um determinado atributo. O conjunto de registros é analisado, sendo que cada um deles já possui a identificação da classe, na intenção de "aprender" como classificar um novo registro (aprendizado supervisionado).

Estimação (Estimation): é semelhante à classificação, no entanto é usada quando o atributo alvo é numérico ao invés de categórico. Basicamente novas observações são baseadas em um modelo de estimação que, por sua vez, advém dos valores de outras variáveis. Por exemplo, estimar a quantia a ser gasta por uma família no período de férias.

Agrupamento (Clustering): objetiva aproximar registros, observações ou classes de objetos similares. Diferentemente das tarefas anteriores, o agrupamento não focaliza o valor de uma variável alvo. Em vez disso, algoritmos de agrupamento procuram segmentos definidos em subgrupos ou grupos relativamente homogêneos, onde a similaridade dos registros dentro do cluster é maximizada e a similaridade fora é minimizada. Alguns exemplos, nas mais variadas áreas, são: processamento de imagens, pesquisa de mercado, reconhecimento de padrões.

Associação (Association): visa descobrir as regras para quantificar a relação entre dois ou mais atributos. As regras de associação são da forma "SE atributo X ENTÃO atributo Y". Por exemplo, de 1000 clientes em um dia de semana a noite, de uma loja, 200 compraram fraldas e, desses 200, 50 compraram cervejas. Logo, a regra de associação poderia ser "Se comprou fraldas, então comprará cervejas".

4.3 Pré-processamento dos dados

O conhecimento obtido após o término oficial de um processo de KDD está intimamente ligado a qualidade dos dados de entrada. Todavia, se os dados presentes em grandes bases de dados e *data warehouses* reais são inconsistentes, incompletos e com ruídos, conforme elenca (HAN; KAMBER, 2006), então o conhecimento será de má qualidade. Nesse sentido, (LA-ROSE, 2005) diz que "para ser útil para fins da mineração de dados, os dados precisam passar por pré-processamento, na forma de limpeza e transformação de dados."

O processo de preparação dos dados para a mineração, denominado de pré-processamento por (HAN; KAMBER, 2006), compreende as principais etapas:

Limpeza dos dados: Dados tendem a ser incompletos e inconsistentes, assim, essa etapa visa a aplicação de métodos como, por exemplo, preenchimento de campos vazios com valores padrão e agrupamento buscando descobrir melhores valores, para a eliminação de problemas.

Integração dos dados: Como o próprio nome sugere, integração compreende a etapa de integrar diferentes fontes de armazenamento de dados em um único repositório consistente. A partir disso, deve-se observar questões de redundância e dependência existentes entre atributos.

Transformação dos dados: Essa etapa merece atenção, pois alguns algoritmos trabalham apenas com valores numéricos e outros com valores categóricos. Sendo assim, é necessária a transformação entre valores categóricos/numéricos de acordo com os objetivos pretendidos. Algumas etapas podem ser citadas para esse fim: suavização de dados discrepantes (*outliers*), agrupamento, generalização, normalização e a criação de novos atributos.

Redução dos dados: Um *data set* pode ser muito grande ocasionando demora e até inviabilidade no processamento e análise de dados. Assim, essa etapa busca diminuir o volume de um *data set* através da remoção de atributos que não denigram a integridade dos dados originais. Com isso, minerar em dados reduzidos seria mais rápido e, ainda, produziria o mesmo resultado.

5 METODOLOGIA

O início do trabalho se dará através do estudo e revisão de literatura de alguns assuntos: técnicas de mineração e pré-processamento de dados. Concomitante a isso, o desenvolvimento de um sistema *web* responsável por receber os dados fornecidos pelos cursistas. Tais dados serão armazenados em um banco de dados relacional, a ser construído. Na sequência, com o banco de dados populado, conhecer o *data set* obtido na intenção de realizar o pré-processamento da melhor forma possível. Por fim, aplicar a técnica de mineração de dados escolhida e averiguar o conhecimento obtido.

6 CRONOGRAMA

	AGO	SET	OUT	NOV	DEZ
Definição do projeto	X				
Construção do banco de dados relacional	X				
Desenvolvimento do sistema web de coleta	X				
Pesquisa sobre as técnicas de mineração de dados	X	X			
Coleta das respostas através do questionário on-line		X			
Identificação e aplicação da técnica de mineração escolhida		X	X		
Revisão de Literatura	X	X	X		
Apresentação do andamento			X		
Elaboração da parte escrita		X	X	X	
Reuniões de acompanhamento	X	X	X	X	X
Apresentação para a banca					X

REFERÊNCIAS

- CABENA, P. et al. **Discovering Data Mining**: from concept to implementation. [S.l.]: Prentice Hall, 1998.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **American Association for Artificial Intelligence**, [S.l.], 1996.
- HAN, J.; KAMBER, M. **Data Mining, Southeast Asia Edition**: concepts and techniques. [S.l.]: Elsevier Science, 2006. (The Morgan Kaufmann Series in Data Management Systems).
- HAND, D.; MANNILA, H.; SMITH, P. **Principles of Data Mining**. [S.l.]: MIT Press, 2001.
- LAROSE, D. **Discovering Knowledge in Data**: an introduction to data mining. [S.l.]: Wiley, 2005.
- LUNARDI, G. M.; LUNARDI, E. M. Gestão Educacional na modalidade EAD: um olhar no processo tecnológico do pró-conselho/ufsm. **Interfaces entre políticas públicas/gestão educacional e formação docente [livro eletrônico]**, [S.l.], 2014.
- MEC. **Sistemas do Ministério da Educação - MEC**. Disponível em <<http://goo.gl/39AY0x>>, Acesso em: 13 ago. 2014.
- REZENDE, S. O. Mineração de Dados. **V Encontro Nacional de Inteligência Artificial**, [S.l.], 2005.
- SICME. **Sistema de Informações dos Conselhos Municipais de Educação**. Disponível em <<http://sicme.mec.gov.br/>>, Acesso em: 12 ago. 2014.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. [S.l.]: Pearson Education India, 2006.