

Rezende [16] explica que, sob a perspectiva da área de Banco de Dados, por exemplo, a Mineração de Dados trata do processo de descoberta de conhecimento armazenado em Base de Dados. De forma complementar, para a área de Aprendizado de Máquina, a Mineração de Dados é apontada como meio de extrair conhecimento potencialmente útil, implícito em dados e que era, no entanto, previamente desconhecido. Na perspectiva da área de Estatística, continua a autora, novas formas de dados podem emergir de resumos de relacionamentos inesperados de conjuntos de dados supervisionados em grandes quantidades.

Para Quoniam et al [15], as ferramentas da Mineração de Dados identificam as correlações existentes nas fontes de dados e, neste sentido, através das técnicas de exploração de dados, extraem informações críticas. Estas, por sua vez, subsidiam o que os autores explanam como “processo decisório de uma organização”. Aplicado ao uso acadêmico e científico de uma informação extraída de bancos de dados relevantes do ponto de vista da pesquisa, tal processo decisório pode ser entendido como o uso do conhecimento na própria pesquisa, ou seja, na maneira de abordagem da ideia que acaba por dar subsídio a defesa de determinada linha de pensamento.

Em um artigo que apresentou o emprego de técnicas de mineração de dados para investigar os temas centrais da pesquisa em design em um periódico da área, Ribeiro et al [18] indicam que utilizaram softwares específicos para proceder a pesquisa. Esta se subdividiu em três principais momentos: a identificação do problema, o pré-processamento e a transformação.

O primeiro momento, identificação do problema, foi relacionado com o volume de informações, quando da grande quantidade de dados não se pôde, ainda, extrair conhecimento. Isso se deu porque o elevado número de informações torna inviável a análise por meio de métodos tradicionais, tais como gráficos e planilhas. Neste caso, a identificação de relações, quando existem, se limita às correlações.

No momento de pré-processamento, Ribeiro et al [18] usaram os procedimentos de organização dos dados coletados. No estudo por eles explanado, as ferramentas dos softwares foram integradas a partir do desenvolvimento de

tabulações e planilhas. A “conversa” entre os softwares foi o primeiro passo para viabilizar a análise das informações.

Para tanto, os dados foram preparados no momento denominado de transformação. Foram decodificados ou transformados de maneira que se viabilizasse o emprego de análises estatísticas ou de extração de padrões. Os agrupamentos e as identificações de correlações deste momento permitiram a definição dos resultados.

De fato, identificar problema, pré-processar e transformar dados são, basicamente, os passos que levam à Mineração de Dados. Rezende et al [17], neste contexto, explanam o processo de Mineração de Dados subdividindo-o em cinco etapas: (i) conhecimento do domínio; (ii) pré-processamento; (iii) extração de padrões; (iv) pós-processamento; (v) utilização do conhecimento.

O (i) conhecimento do domínio refere-se a seleção de dados em determinada base, considerando os objetivos da exploração. É a identificação do problema, quando devem ser respondidas questões importantes em relação às metas do processo, às restrições, aos critérios e aos objetivos com a extração do conhecimento. Esta etapa é fundamento para as seguintes.

O (ii) pré-processamento é o tratamento que os dados recebem a fim de serem utilizados nas etapas posteriores. Os dados são higienizados, ou seja, são tratados, limpos e sofrem redução de volume como meio de preparação para a extração de padrões. As transformações nos dados são efetivadas por meio da extração e integração, a organização e a limpeza, a seleção e a redução.

A (iii) extração de padrões tem o objetivo de encontrar modelos a partir dos dados. Visa a obtenção de resultados mais adequados aos objetivos a que se propõe o levantamento, ou seja, ao cumprimento das metas preestabelecidas. Nesta etapa podem ser necessários ajustes que permitam maior precisão e melhor compreensão do conhecimento extraído. Para o devido ajuste de parâmetros, pode ser necessário executar esta etapa mais de uma vez.

O (iv) pós-processamento é a avaliação quanto a qualidade dos dados selecionados. A passagem por esta etapa permitirá que os dados extraídos sejam considerados para a resolução de