

investigação, o objetivo é, conforme explicam Prodanov e Freitas [13], a geração de conhecimentos que envolvem interesses para a solução de problemas específicos em uma aplicação prática.

O estudo de caso está sendo nomeado, nesta pesquisa, remetendo a aplicação prática. Nele, um estudo real, ou seja, o uso de dados de uma pesquisa científica em desenvolvimento, foi usado para testar o método. Todo o processo foi descrito na forma de passos constituintes do objeto de estudo – o método aqui apresentado.

### CONCEITOS DE MINERAÇÃO DE DADOS

Uma revisão sistemática, segundo Flick [10], segue critérios específicos e um foco estreito em torno de um tema geral de pesquisa. As direções da revisão sistemática são os próprios documentos de pesquisa. Sua utilização, explica o autor, implica em pré-definir critérios a fim de torná-la replicável e, em si mesma, passível de avaliação.

É, portanto, indispensável que, ao realizar uma coleta de dados, estes estejam primeiramente organizados e, preferencialmente, esquematizados de modo claro. A organização e a clareza darão suporte para que a pesquisa seja replicada – e assim confirmada ou contestada – e avaliada em termos de processo ou de qualidade.

Faz parte do processo de pesquisa científica uma revisão da literatura. Esta consiste em fazer uma verificação em torno do que já foi escrito e publicado sobre o assunto de interesse, quem escreveu e quando, se há lacunas e/ou oportunidades de abordagens que evitem a duplicidade de estudos no mesmo enfoque. [12]

A revisão da literatura apoiada por um processo sistemático torna-se, então, relevante tanto para a localização de informações científicas quanto para a exploração consciente, não aleatória, dos dados. Ela determina o “estado da arte” e a “definição de contornos mais precisos do problema a ser estudado” [10].

Por outro lado, a pergunta sobre “por onde começar?” faz parte desse processo. Ao se deparar com extensos bancos de dados e acesso facilitado às informações científicas disponibilizadas em variados meios, o pesquisador corre o risco de “se perder” no objetivo da pesquisa e, com isso, desfocar a busca. Com a não organização da busca, dados

relevantes podem deixar de ser acessados ou, ainda, o pesquisador coleta um número tão grande de dados que não tem condições de analisá-los devidamente.

Neste sentido, a Mineração de Dados pode contribuir munindo o pesquisador de passos e ferramentas que auxiliam na organização da busca e, com isso, em melhor uso da informação coletada. A Mineração de Dados ou *Data Mining*, conforme Rezende [16], é o processo de extração de conhecimento a partir de grandes volumes de dados. Mais especificamente, explica a autora, visa “a obtenção de conhecimento útil e interessante, a partir de dados, para a utilização em um processo de tomada de decisão”.

*Data Mining* é uma das técnicas utilizadas para realização de KDD – *Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Bases de dados – que é o processo de “garimparem” de informações relevantes que tem crescido e se complexizado com o aumento da capacidade de armazenamento em bancos de dados e o avanço das tecnologias de informação e comunicação.[15]

Usando as técnicas de Mineração de Dados, diferentes tarefas podem ser executadas sobre as bases que guardam e compartilham informações científicas. Tais técnicas são úteis quando busca-se a geração de conhecimento emergido da associação por meio de aplicação de regras, da classificação dos dados coletados, da clusterização ou agrupamentos dos registros pela similaridade dos atributos ou características e, ainda, da visualização de anomalias com o objetivo de detectar desvios do padrão. [14]

No contexto que contempla o grande volume de dados armazenados – e hoje disponibilizados, de forma acessível, por instituições em sua maioria acadêmicas – e que embutem, muitos deles, informações realmente novas e interessantes, a Mineração de Dados se constituiu como forma interdisciplinar. Rezende [16] aponta que Mineração de Dados tem estreitas relações com as contribuições oferecidas por diversas áreas como Banco de Dados, Aprendizado de Máquina, Estatística, Recuperação de Informação, Computação Paralela e Distribuída.

Por esta interdisciplinaridade, a Mineração de Dados é explicada por diferentes perspectivas que acordam com a sua área de abordagem. Citando autores de cada uma dessas áreas,