

2 | MINERAÇÃO DE DADOS EDUCACIONAIS

Mineração de dados (MD) é uma área que explora grandes volumes de dados em busca de padrões. Para isso, existem algoritmos de machine Learning, como árvores de decisão, que são capazes de fazer com que o computador aprenda usando dados de eventos passados. Fayyad (1996) propôs um processo para transformar um conjunto de dados em padrões (conhecimento), que é conhecido como processo de descoberta de conhecimento de bases de dados (Knowledge Discovery in Databases), composto de 5 etapas, conforme é ilustrado na Figura 1.

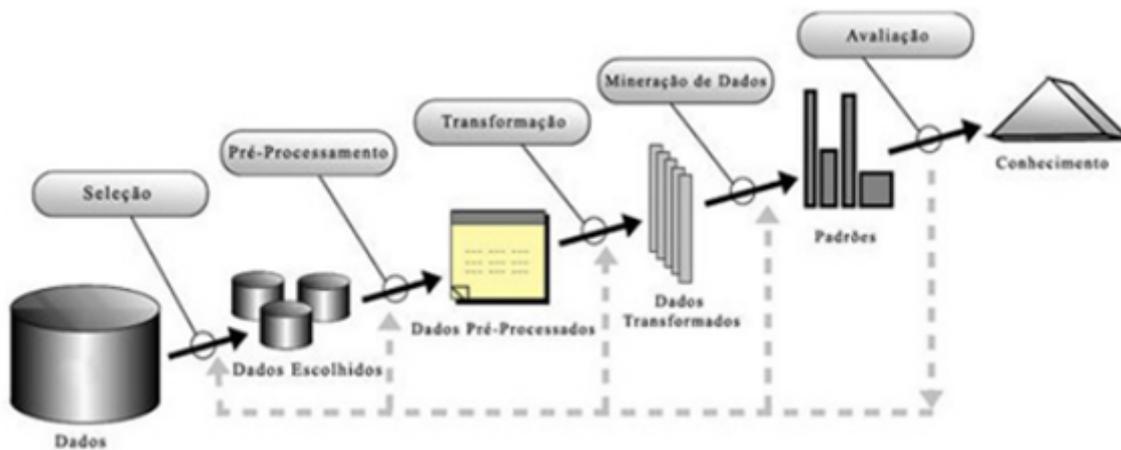


Figura1. Etapas do Processo KDD.

Fonte: (FAYYAD, 1996).

Na etapa de seleção, o objetivo é definir/escolher dados (atributos/características) a serem minerados a partir de dados brutos, podendo ser extraídos de diversas fontes, tais como diários escolares, planilhas eletrônicas, data warehouses, entre outros. A etapa de pré-processamento envolve operações como tratar a falta de dados em alguns atributos, limpeza de dados, redução da quantidade de atributos (características), preenchimento ou eliminação de valores nulos e remoção de dados duplicados. A etapa de transformação, basicamente, se resume em formatar dados para serem interpretados pelos algoritmos de aprendizado, ou seja, alterar o formato do arquivo de dados. A etapa de mineração de dados caracteriza-se pela aplicação de um ou mais algoritmos para extrair conhecimento implícito em padrões. E a etapa de avaliação consiste em interpretar e validar o conhecimento descoberto para processos de tomada de decisão, sendo necessário, o processo pode recomeçar por uma etapa anterior.

Segundo Baker (2009), grande parte dos métodos utilizados em Mineração de Dados Educacionais (Educational Data Mining- EDM) são oriundos da própria mineração de dados, adaptados às necessidades e particularidades da área da Educação. Em Baker (2009), é possível encontrar uma

lista de tarefas de aprendizado que podem ser realizadas com dados educacionais, sendo que cada uma possui um objetivo específico. Neste trabalho, a tarefa de aprendizado escolhida foi classificação porque busca encontrar características de alunos que evadiram analisando os fatores que contribuíram.

Revisão #1

Criado 7 outubro 2021 14:38:58 por Valerio Augusto Lopes Passos

Atualizado 7 outubro 2021 14:39:26 por Valerio Augusto Lopes Passos