1-

A Mineração de dados (MD) compreende um processo orientado a extração de conhecimento implícito, previamente desconhecido, e útil, a partir de grandes repositórios de dados (WITTEN; FRANK, 2005). Essa área é formada por ferramentas e técnicas que utilizam [algoritmos](https://pt.wikipedia.org/wiki/Algoritmo) de aprendizagem ou classificação para explorar um conjunto de dados, extraindo ou evidenciando padrões nestes dados e auxiliando na descoberta de conhecimento. (J. HAN, M. KAMBER, J.PEI, 2011).

Segundo Fayyad (et al., 1996), a mineração de dados é “um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados”.

Existem muitas ferramentas utilizadas para executar algoritmos de mineração, e as mesmas contribuem para que os especialistas concentrem esforços apenas em partes mais significativas dos dados (SELIYA, KHOSHGOFTAAR, 2007). Algumas ferramentas a serem citadas, são: Weka,Oracle Data Mining (ODM), SAS Text Miner, KNIME, dentre outras (CAMILO, SILVA, 2009).

Algoritmos utilizados pela MD, em diferentes tarefas: Apriori, k-means, k-medoids, MTS, CART, Algoritmo Genético Simples,CHC, Redes de Kohonen, Rede Bam, dentre outros.

2-

A Mineração de Dados é classificada pela sua capacidade em realizar determinadas tarefas (LAROSE, 2005), tais como:

**Descrição:** pode ser definida como a tarefa usada para descrever padrões e tendências dos dados. Esta tarefa oferece uma possível interpretação para os resultados obtidos, sendo muito utilizada em conjunto com técnicas de análise exploratória de dados, para comprovar a influência de certas variáveis no resultado obtido.

**Classificação:** visa identificar a qual classe um determinado registro pertence. O modelo deve analisar os registros fornecidos, com cada registro contendo a indicação da classe pertencente, com o intuito de “aprender” como classificar um novo registro (HAN, KAMBER, 2006). Esta tarefa pode ser usada em diversos casos, como por exemplo:

• Determinar quando uma transação de cartão de crédito pode ser uma fraude;

• Diagnosticar onde uma determinada doença pode estar presente;

• Identificar quando uma pessoa pode ser uma ameaça para a segurança.

**Estimação:** esta tarefa é usada em registros identificados por valores numéricos (ao invés de categóricos). Desse modo, é possível estimar o valor de uma variável analisando-se os valores das demais. (GONZÁLEZ-ARANDA, MENASALVAS, RUIZ, SEGOVIA, 2008) Por exemplo: um conjunto de registros contendo os valores mensais gastos por diversos tipos de consumidores e de acordo com os hábitos de cada um. Após ter analisado os dados, o modelo é capaz de dizer qual será o valor gasto por um novo consumidor.

**Predição:** esta tarefa visa descobrir o valor futuro de um determinado atributo.

Alguns exemplos de predição:

• Predizer o valor de uma ação três meses adiante;

• Predizer o percentual que será aumentado de tráfego na rede se a velocidade aumentar;

**Agrupamento:** têm o objetivo de identificar e aproximar os registros similares. Um cluster é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. O agrupamento não necessita que os registros sejam previamente categorizados. As aplicações das tarefas de agrupamento são as mais variadas possíveis: pesquisa de mercado, reconhecimento de padrões, processamento de imagens, análise de dados, entre outras (OLIVEIRA, CARVALHO, 2008). Geralmente a tarefa de agrupamento é combinada com outras tarefas, além de serem usadas na fase de preparação dos dados.

**Associação:** consiste em identificar quais atributos estão relacionados. Apresentam a forma: SE X ENTÃO Y. É uma das tarefas que apresenta bons resultados obtidos, principalmente na análise da "Cestas de Compras", em que identifica-se quais produtos são levados juntos pelos consumidores. Alguns exemplos:

• Determinar os casos onde um novo medicamento pode apresentar efeitos colaterais;

• Identificar os usuários de planos que respondem bem a oferta de novos serviços.

**Massa de dados**

Pode ser definida como o conjunto de dados, ou ainda aglomerado de dados utilizados por uma aplicação. Massas de dados geralmente são utilizadas para testar paginação de grids, popular bancos de dados para realizar testes de performance, fazer testes unitários, exibir retorno em API que está sendo construída, dentre outros (Dias, 2002).

**Referências**

WITTEN, I. H.; FRANK, E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

J. HAN; M. KAMBER; J. PEI – Data Mining: Concepts and Techniques (2011).

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in knowledge discovery and data mining. In: FAYYAD, U. M. et al. (Ed.). Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. cap. From Data Mining to Knowledge Discovery: An Overview, p. 1–34. ISBN 0-262-56097-6

CIOS, K. J; PEDRYCZ, W; SWINIARSKI, R. W; KURGAN, L. A. Data Mining - A Knowledge Discovery Approach. Springer, 2007.

SELIYA, N; KHOSHGOFTAAR, T. M. Software Quality Modeling With Limited Apriori Defect Data, chapter Chapter 1, p. 1–16. Idea Group Publishing, 2007.

Wang, J, editor. Encyclopedia of Data Warehousing and Mining. Idea Group Reference, 2005.

LAROSE, D. T. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley and Sons, Inc, 2005.

OLIVEIRA, R. R; CARVALHO, C. L. Algoritmos de agrupamento e suas aplicações. Technical report, Universidade Federal de Goiás, 2008.

CAMILO, C.; SIlVA, C. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. Technical report, Universidade Federal de Goiás, 2009.

HAN, J; KAMBER, M. Data Mining: Concepts and Techniques. Elsevier, 2006

GONZÁLEZ-ARANDA, P; MENASALVAS, E; RUIZ, S. M. C; SEGOVIA, J. Towards a methodology for data mining project development: The importance of abstraction. In: STUDIES IN COMPUTATIONAL INTELLIGENCE, p. 165U178. Springer-Verlag, ˝ 2008.

DIAS, M. M. Parâmetros na escolha de técnicas e ferramentas de mineração de dados. Departamento de Informática, Universidade Estadual de Maringá, 2002.

DANNY. *37 Mind Blowing YouTube Facts, Figures and Statistics – 2018.* Abril de 2018, sl. Disponível em <<https://merchdope.com/youtube-statistics/>>. Acesso em 8 Abr. 2018.