## What is Data Mining?

Data mining is essentially the analysis of large datasets, which involves the process of extracting and discovering anomalies, patterns, correlations and other valuable information from data points within the large dataset. Data mining is also known as “knowledge discovery in databases” and the concept itself has been around for over a century but the term “data mining” was only coined in the 1990’s. In today’s world, data mining is a field of computer science which combines the fields of statistics, artificial intelligence, neural networking and machine learning. Its purpose is to identify trends and extract information strictly in large datasets which would have otherwise remained hidden or overlooked.

In recent years, data mining has only gotten increasingly popular and prevalent alongside the growing prevalence of big data. This is due to datasets being too large and complex to analyze manually and therefore rely on machine learning algorithms in order to help analyze and extract meaningful information from them. This is useful as it significantly speeds up the process in which meaningful insights can be extracted from ever-growing datasets, especially as technologies improve the rate at which data and information is generated.

The process of data mining is generally done in standardized steps to ensure the quality of the extracted data is meaningful and accurate, choosing carefully and wisely the input datasets to be used for analysis, as bad data will only yield poor results:

1. **Business Understanding** - Understanding the objectives and parameters of the data mining project and setting what is appropriate in order to initiate and run a fruitful project.
2. **Data Understanding** - Determining the types of data which will be involved and carefully considering the source of required data.
3. **Data Preparation** - Preparing the data to be analyzed, removing and fixing any issues such as duplicate or missing data. Also optionally reducing the depth or complexity of the dataset if deemed to be too computationally intensive.
4. **Modeling** - This is the most interesting step of the process where algorithms will be used to analyze and identify any meaningful relationships or deviations in the data.
5. **Evaluation** - Reviewing and interpreting the information gathered from the modeling phase and determining whether or not it will be useful and whether or not the goal of the project has been achieved. Multiple different algorithms may be used and determining which one is best suited or may have achieved the best results.
6. **Deployment** - Presenting the results and implementing solutions based on the data gathered.

Data mining can be applied and is useful in various industries such as Tech, Business, Education, Manufacturing, and many more. In tech, it can be used for analytics and to predict future outcomes and trends to better prepare for changes in the future. In business, it can help in sales forecasting and to help in adapting marketing strategies to better suit their target audiences or to expand their reach. As for education, it can be used in order to identify behavioural trends in students which could help adapt teaching techniques and develop new methods in order to adjust and improve learning to be better suited to more students. There are many possible applications to data mining and it will only ever increase as the amount of scenarios where big data is applied increases as well.

Resources:

<https://www.sas.com/en_us/insights/analytics/data-mining.html>

<https://www.ibm.com/cloud/learn/data-mining>

<https://www.britannica.com/technology/data-mining>

<https://www.talend.com/resources/what-is-data-mining/>

<https://www.sas.com/en_us/industry/education.html>