**Data Mining**

Data mining is the process through which we go through data in order to sort and identify patterns and relationships, this analysis helps solving large scale industrial problems. The inferred knowledge based on the tools and techniques used in the data mining helps companies predict future trends making informed decisions.

Data mining, a key part of data analytics and one of the core disciplines in data science, uses advanced analytics techniques to extract useful information in data sets. Knowledge discovery database (KDD) and data mining are sometimes used interchangeably because KDD at more detailed level includes a step which is data mining.

***Significance of Data Mining***

The information Data mining generates can be used Business Intelligence (BI) and advanced analytics applications that involve analysis of historical data, as well as real-time analytics applications that examine streaming data as it's created or collected.

Effective data mining aids in various aspects of planning business strategies and managing operations. That includes customer-facing functions such as marketing, advertising, sales and customer support, plus manufacturing, supply chain management, finance and HR. Data mining supports fraud detection, risk management, cybersecurity planning and many other critical business use cases. It also plays an important role in healthcare, government, scientific research, mathematics, sports and more.

***Data Mining Methodology***

Relevant data identified and assembled.

Data Sourcing

Data exploration, profiling and opre-processing and finally data quality control.

Data Exploration/Gathering

Users create model, test it, and then evaluate.

Modelling

Taking an actions based on the results from the models.

Deploying Models

***Data mining software and tools***

Key features provided by data mining software include data preparation capabilities, built-in algorithms, [predictive modeling](https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling) support, a GUI-based development environment, and tools for deploying models and scoring how they perform.

Vendors that offer tools for data mining include Alteryx, AWS, Databricks, Dataiku, DataRobot, Google, H2O.ai, IBM, Knime, Microsoft, Oracle, RapidMiner, SAP, SAS Institute and Tibco Software, among others.

A variety of free open source technologies can also be used to mine data, including DataMelt, Elki, Orange, Rattle, scikit-learn and Weka. Some software vendors provide open-source options, too. For example, Knime combines an open-source analytics platform with commercial software for managing data science applications, while companies such as Dataiku and H2O.ai offer free versions of their tools.