Data Mining is a process of finding potentially useful patterns from huge data sets. It is a multi-disciplinary skill that uses machine learning, statistics, and AI to extract information to evaluate future events probability. The insights derived from Data Mining are used for marketing, fraud detection, scientific discovery, etc.

Data Mining is all about discovering hidden, unsuspected, and previously unknown yet valid relationships amongst the data. Data mining is also called Knowledge Discovery in Data (KDD), Knowledge extraction, data/pattern analysis, information harvesting, etc.

Types of Data

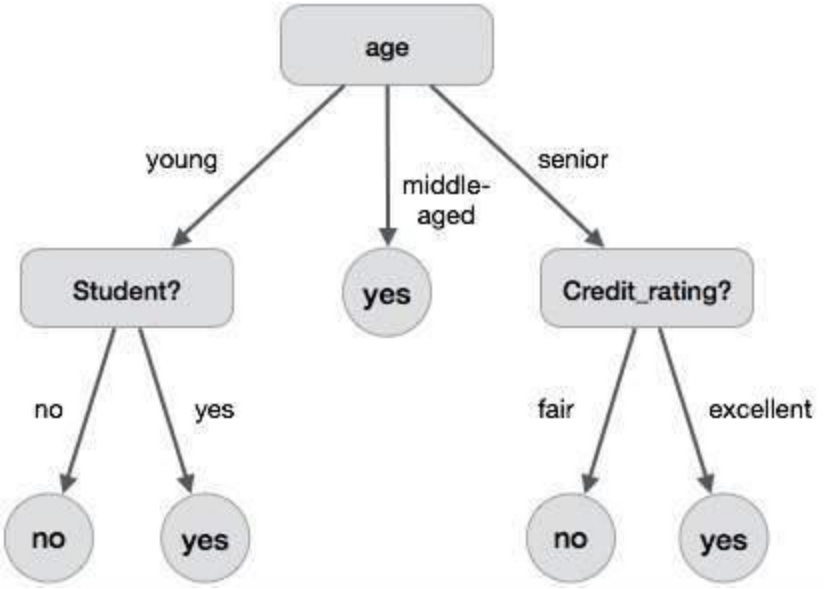
Data mining can be performed on following types of data

- Relational databases
- Data warehouses
- Advanced DB and information repositories
- Object-oriented and object-relational databases
- Transactional and Spatial databases
- Heterogeneous and legacy databases
- Multimedia and streaming database
- Text databases
- Text mining and Web mining

Data Mining Techniques

Decision Trees: A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

The following decision tree is for the concept buy_computer that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents a test on an attribute. Each leaf node represents a class.



The benefits of having a decision tree are as follows −

- It does not require any domain knowledge.
- It is easy to comprehend.
- The learning and classification steps of a decision tree are simple and fast.

Neural Networks: **Neural networks are used for effective data mining in order to turn raw data into useful information. Neural networks look for patterns in large batches of data, allowing businesses to learn more about their customers which directs their marketing strategies, increase sales and lowers costs.**

Neural networks mine data in areas such as bioinformatics, banking, and retail. Using neural networks, data warehousing organisations can harvest information from datasets to help users make more informed decisions through neural network's ability to handle complex relationships, cross-pollination of data, and machine learning. Neural networks and AI technologies can carry out many business purposes with unstructured data, from tracking and documenting real-time communications, to finding new customers that automate follow-ups and flag warm leads

Nearest Neighbor & Clustering :

Nearest Neighbor (also known as Collaborative Filtering or Instance-based Learning) is a useful data mining technique that allows you to use your past data instances, with known output values, to predict an unknown output value of a new data instance. So, at this point, this description should sound similar to both regression and classification. How is this different from those two? Well, first off, remember that regression can only be used for numerical outputs. That differentiates it from Nearest Neighbor immediately.

- Genetic Algorithms
- Rule Introduction
- Classification: This analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes.
- Clustering: Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data.
- Regression: Regression analysis is the data mining method of identifying and analyzing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.
- Association Rules: This data mining technique helps to find the association between two or more Items. It discovers a hidden pattern in the data set.
- Outer detection: This type of data mining technique refers to observation of data items in the dataset which do not match an expected pattern or expected behavior. This technique can be used in a variety of domains, such as intrusion, detection, fraud or fault detection, etc. Outer detection is also called Outlier Analysis or Outlier mining.
- Sequential Patterns: This data mining technique helps to discover or identify similar patterns or trends in transaction data for certain period.
- Prediction: Prediction has used a combination of the other techniques of data mining like trends, sequential patterns, clustering, classification, etc. It analyzes past events or instances in a right sequence for predicting a future event.