

# [Data frequently used techniques are clustering, classification,](https://assignbuster.com/data-frequently-used-techniques-are-clustering-classification/)

Data mining is the process of analyzingdata from different perspectives and summarizing it into useful information. Data mining software is one of the numbers of analytical tools for analyzingdata. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. The aim of datamining is to discover structure inside unstructured data, extract meaning fromnoisy data, discover patterns in apparently random data, and use all thisinformation to better understand trends, patterns, correlation in existing data. One of the most important tasks inData Mining is to select the correct data mining technique.

Data miningtechnique has to be chosen based on the type of business and the type ofproblem your business faces. A generalized approach has to be used to improvethe accuracy and cost effectiveness of using data mining techniques. Morefrequently used techniques are clustering, classification, association, decision tree, neural network. In this paper we discus about the clusteringtechniques briefly.

Data mining is a technology that consiststraditional data analysis methods with sophisticated algorithms for processinglarge volumes of data through which useful information and knowledge areextracted. Data mining is the process of extracting or mining Information fromhuge amounts of data. It is also known as knowledge mining from data. DataMining is the method of extracting information from large data sets through theuse of algorithms and techniques drawn from the field of Statistics, MachineLearning and Data Base Management Systems. Clustering is unsupervised learning itdeals with finding structure in a collection of unlabeled data. These modelsare sometimes called descriptive model.

Clustering is one of the most commonuntested data mining methods that explore the hidden structures in a dataset. Data mining approaches are useful bioinformatics. In bioinformatics Geneexpression, DNA sequence, RNA sequence and protein sequence clusters is knownas clustering The data objects which are more similar in nature are clusteredinto same cluster while the data objects which are highly dissimilar lie inseparate clusters. Fundamentally, clustering is an unsupervised classificationwhere the class label is unknown.

There are several major data mining techniques havebeen developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns etc.,