

A brief explanation of the mentioned concepts will be discussed hereafter:

1. **Clustering:** a type of Machine Learning tasks in which the model attempts to detect patterns in a given data set and define classes of labels by itself based on certain predefined parameters. Algorithms belonging to this type usually work with unlabeled data. Some typical algorithms for this task are K-Means and DBSCAN
2. **Visualization:** algorithms in this type are generally for working with complex and unlabeled data. The algorithm attempts to represent the data set in a 2D or 3D space, which are most comprehensible to the human brains, thus helps data analysts detect patterns of the data points without changing drastically the structure of the data set.
3. **Dimensionality reduction:** it is very common that the collected data is in a very complex format and not all features are useful for every ML model. This is where dimensionality reduction comes in to simplify data structure as well as save computational resource. It involves a number of techniques to select only the most relevant features or merge them together to reduce the amount of memory being occupied by unnecessary data as well as the computing power being spent on extraneous operations.
4. **Feature extraction:** this is one of the dimensionality reduction techniques, which selects only representative features of the data set or merges those with high correlation together. This helps freeing up storage space and reducing the computing power needed to process data or train Machine Learning models while still keeps an accurate presentation of the original data set.
5. **Anomaly detection:** a type of Machine Learning tasks which the model attempts to detect unusual patterns/data points and report or make decisions upon detection. This is achieved with the aid of different mathematical operations and algorithms and a set of data samples which are considered “normal”.
6. **Association rule learning:** A Machine Learning method which helps revealing interesting and most likely hidden relations between variables in a data set. This is a useful tool when it comes to feature extraction and selection as the correlations among variables are evaluated and measured against certain metrics, which plays an important role in selecting training data to avoid overfitting or underfitting.