PNT2022TMID21280_Efficient Water Quality Analysis & Prediction using Machine Learning.

Supervised and Unsupervised Learning:

Supervised machine learning requires labelled input and output data during the training phase of the machine learning lifecycle. This training data is often labelled by a data scientist in the preparation phase, before being used to train and test the model. Once the model has learned the relationship between the input and output data, it can be used to classify new and unseen datasets and predict outcomes.

The reason it is called supervised machine learning is because at least part of this approach requires human oversight.

Unsupervised machine learning is the training of models on raw and unlabelled training data. It is often used to identify patterns and trends in raw datasets, or to cluster similar data into a specific number of groups. It's also often an approach used in the early exploratory phase to better understand the datasets.

As the name suggests, unsupervised machine learning is more of a hands-off approach compared to supervised machine learning. A human will set model hyperparameters such as the number of cluster points, but the model will process huge arrays of data effectively and without human oversight.

A main difference between supervised vs unsupervised learning is the problems the final models are deployed to solve. Both types of machine learning model learn from training data, but the strengths of each approach lie in different applications. Supervised machine learning will learn the relationship between input and output through labelled training data, so is used to classify new data using these learned patterns or in predicting outputs.

Unsupervised machine learning on the other hand is useful in finding underlying patterns and relationships within unlabelled, raw data. This makes it particularly useful for exploratory data analysis, segmenting or clustering of datasets, or projects to understand how data features connect to other features for automated recommendation systems.

Examples of supervised machine learning include:

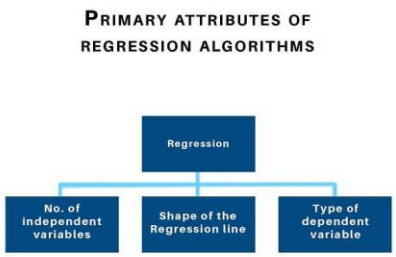
- Classification, identifying input data as part of a learned group.
- Regression, predicting outcomes from continuously changing data.

Examples of unsupervised machine learning include:

- Clustering, grouping together data points with similar data.
- Association, understanding how certain data features connect with other features.

Regression:

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).



Data Visualization:

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

- infographics
- bubble clouds
- bullet graphs
- heat maps
- fever charts
- time series charts