tinubu®

Glossary of Artificial Intelligence & Quantum Computing

by Thomas Frossard

Tabular data
It refers to data that is organized in a table with rows and columns.

Time-series
In mathematics[1], it's a series of data points[2] indexed in time order.

Unstructured data
Information in many different forms that doesn't follow conventional data models, making it difficult to store and manage in a mainstream relational database[3].

Data sourcing
The process by which companies extract and integrate data from multiple internal and external sources.

Data engineering
The process of preparing data for analytical or operational uses. A few steps of Data engineering are extracting the data, storing the data, cleaning the data, and giving access to the data to a third party.

Imbalanced data set
A dataset where the distribution of the different classes is very disparate. Training an algorithm without rebalancing the dataset would introduce biases that would reduce its performance.

Synthetic data
A piece of data that doesn't come from real recordings but is rather created, either from real data or through a simulation. Synthetic data are often generated when the amount of real-world data is insufficient.

Statistics
A branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data. The mathematical theories behind statistics rely heavily on differential and integral calculus, linear algebra, and probability theory.

Machine Learning
A field of artificial intelligence based on mathematical and statistical approaches to give computers the ability to "learn" from data. It regroups Natural Language Processing (NLP), Ensemble methods (random forest), Deep learning, Clustering, etc.

Deep Learning
A part of the family of machine learning based on artificial neural networks, learning can be supervised semi-supervised or unsupervised. A few examples are Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Fully Connected Neural Network, etc.

Quantum computing
A new computation paradigm supposed to bring exponential speedups by leveraging the properties of quantum mechanics such as superposition and entanglement.

Classification
A particular set of machine learning problems that aims to identify which of a set of classes an observation belongs to.

Regression
A particular set of machine learning problems where the goal is to estimate a numerical output from a set of input features.

Optimization
The process of making the performances of a model or algorithm as good as possible.

Overfitting
When a model learns the specificities of a particular dataset rather than the general underlying relations. An overfitted model would fail on additional data.

Performance metrics
There are various ways of measuring the performances of a classification model.

Among them ranging from 0 (worst) to 1 (best).

- **Accuracy**: the most basic one, the part of the data that has been correctly classified by the model.

- **ROC AUC**: provides an aggregate measure of a model's performance by tweaking its classification threshold.

- **F1-score**: especially used when dealing with imbalanced data where indicators such as accuracy can be misleading.

---

1 https://en.wikipedia.org/wiki/Mathematics
2 https://en.wikipedia.org/wiki/Data_point
3 https://searchsqlserver.techtarget.com/definition/database