

Project data and analytics glossary

This project data and analytics glossary assists project professionals by defining keywords used in project data analytics. This is best read alongside APM's pathfinder report on *Project data analytics: The state of the art and science* which should provide useful background context.

Keyword	Definition
Algorithm	A mathematical formula or statistical process used to perform analysis of data.
Application program interface (API)	Set of programming standards and instructions for accessing or building web-based software applications.
Artificial intelligence (AI)	Study of 'intelligent agents,' autonomous non-human entities that can take in information from their environment and act upon their environment in a way that enables them to succeed in their goals. Intelligent agents need to have mastered machine learning and aspects of predictive data analytics in order to be able to do this. In a project context, some people have speculated that an intelligent agent could enhance or change the roles and status of many project professionals.
AI sub areas	<p>Can be broken down into three sub areas:</p> <ul style="list-style-type: none"> ■ Artificial Narrow Intelligence (ANI) also known as "Weak" AI is the AI that exists in our world today. Narrow AI is AI that is programmed to perform a single task e.g. checking the weather, playing chess or analysing raw data to write journalistic reports. ■ Artificial General Intelligence or "Strong" AI refers to machines that exhibit human intelligence. In other words, AGI can successfully perform any intellectual task that a human being can. Often portrayed in sci-fi movies in which humans interact with machines and operating systems that are conscious, sentient, and driven by emotion and self-awareness. ■ Artificial Super Intelligence (ASI) will surpass human intelligence in all aspects — from creativity, to general wisdom, to problem-solving. Machines will be capable of exhibiting intelligence that we haven't seen in any humans.
Behavioural analytics	Use of data on a person or object's behaviour to make predictions on how it might change in the future or determining variables which affect it, so more favourable or efficient outcomes might be achieved.
Big data	This refers to extremely large bodies of data (or datasets). In project terms, this often refers to the historic 'data plumes' of legacy that are created from the use of project control or enterprise management systems. Project data analytics (both predictive and descriptive) uses Big Data in its activities.
Business intelligence (BI)	The general term used for the identification, extraction, and analysis of data.

Clustering	Clustering techniques attempt to collect and categorise sets of points into groups that are "sufficiently similar," or "close" to one another. "Close" varies depending on how you choose to measure distance. Complexity increases as the more features are added to a problem space.
Data engineering	The collection and storage of data which allows for batch or real-time processing for data scientists to query.
Data ethics	The practices and policies of an organisation that ensure that data is used not only in a way that is compliant with regulations, but in an ethical manner that would stand up to external scrutiny. For example, how is the data used and for what purpose.
Data governance	A set of processes or rules that ensure data integrity and that data management best practices are met.
Data institution	Data institutions are organisations that steward data by governing access to that data on behalf of a community of other organisations or individuals for the long-term.
Data mining	Finding meaningful patterns and deriving insights in large sets of data using sophisticated pattern recognition techniques. To derive meaningful patterns, data miners use statistics, machine learning algorithms and artificial intelligence.
Data science	Field that works with and analyses large amounts of data to provide meaningful information that can be used to make decisions and solve problems. Data science includes work in computation, statistics, analytics, data mining, and programming.
Data steward	<p>This is a concept that arises out of data governance. It recognises that accountability for things like data quality, metadata and the implementation of data policies needs to be devolved to business departments and often locations. A data steward is the person within a particular part of an organisation who is responsible, on a day-to-day basis, for ensuring that their data is fit for purpose and that their area adheres to data policies and guidelines.</p> <p>Data stewards will typically have a reporting relationship to a data owner, who focuses on similar areas but from more of a strategic perspective.</p>
Data trust	This is one form of data institution which involves taking the concept of a legal trust and applying this to data this can include organisations and individuals. Data trusts provide a legal structure that provides independent stewardship of some data for the benefit of a collection of organisations or individuals for a variety of reasons. Typically a data steward governs and oversees the data is used for the benefit of members of the trust to ensure impartiality.
Data visualisation	Any attempt to make data more easily digestible by rendering it in a visual context. Data visualisation includes charting, graphing, infographics etc.

Decision trees	Basic decision-making structure which can be used by a computer to understand and classify information. By asking a series of questions about each data item fed into them, outputs are channelled along different branches leading to different outcomes, typically labelling or classification of the piece of data.
Deep learning	A branch of machine learning that attempts to mirror the neurons and neural networks associated with thinking in human beings. Examples include speech recognition, translation, and image recognition software.
Descriptive analytics	Presenting data in the most effective format.
IoT (Internet of Things)	<p>The network of physical objects or "things" embedded with electronics, software, sensors and connectivity to enable it to achieve greater value and service by exchanging data with the manufacturer, operator and/or other connected devices.</p> <p>Each thing is uniquely identifiable through its embedded computing system but is able to interoperate within the existing internet infrastructure.</p>
Machine learning	This is the name given to computer algorithms that 'learn from doing'. In project terms, machine learning has, at its centre, algorithms that are used to spot patterns between some characteristic of projects or programmes and some aspect of project performance. This process gets more accurate the more it is used. Machine learning is a fundamental part of predictive project data analytics
Metadata	Data that describes other data. Meta is a prefix that - in most information technology usages - means "an underlying definition or description." Metadata summarises basic information about data, which can make finding and working with particular instances of data easier. For example, blogs, photos, Word documents etc.
Network analysis	Network analysis applied to the study of the social agents responsible for scientific publications allows us to identify the number of members in the network, the intensity of the relationship between them and the most relevant members of the network.
Network analytics	The science of describing and, especially, visualising the connections among objects. The objects might be human, biological or physical.
Predictive analytics	Using past data to predict future performance.
Predictive modelling	The process of developing a model that will most likely predict a trend or outcome.
Project data analytics	Project data analytics, at its simplest, is the use of past and current project data to enable effective decisions on project delivery.
Python	An open source programming language used by data scientists amongst others.

Quantile	A group of objects which have been classified according to similar characteristics, and then distributed evenly between several such groups. These are distinguished as "quartile" if there are four such groups, "quintile" if there are five such groups, etc. The "first quartile" would refer to the top quarter of entries in a list which has been split into four equal groups.
R	An open source programming language used for statistical computing and graphics.
Structured data	Data that is organised according to a predetermined structure (rows and columns, as an example).
Supervised machine learning	With supervised learning techniques, the data scientist gives the computer a well-defined set of data. All the columns are labelled, and the computer knows exactly what it's looking for. It's like a professor handing you a syllabus and telling you what to expect on the final.
Transactional data	Data that relates to the conducting of business, such as accounts payable and receivable data or product shipments data.
Unstructured data	Data that has no identifiable structure, i.e. email message text, social media posts, audio files such as recorded human speech, music, etc.
Unsupervised machine learning	In unsupervised learning techniques, the computer builds its own understanding of a set of unlabelled data. Unsupervised machine learning techniques look for patterns within data, and often deal with classifying items based on shared traits.

Sources

Bernard Marr & Co

<https://www.bernardmarr.com/default.asp?contentID=1446>

Data Quest

<https://www.dataquest.io/blog/data-science-glossary/>

Lean Methods Group

<https://leanmethods.com/resources/articles/data-analytics-glossary/>

Royal Society

<https://royalsociety.org/topics-policy/data-and-ai/>

The Alan Turing Institute

<https://www.turing.ac.uk/>

The Open Data Institute

<https://theodi.org>