

**Lesson 3.2 Discovering Knowledge from Data Key Terms**

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **Abstraction in Models** | Abstraction is generalization with loss of detail. A model abstracts many details, often by reducing several factors into a single parameter. The use of a single number for the reflectivity of clouds in a climate model is an abstraction. |
| **Comma Separated Value File (CSV)** | The most widely readable form of data. A file format in which each line contains multiple values separated by a comma. There are many CSV standards. Most of them allow string values which can contain commas inside a quote. |
| **Confidence Interval** | A range of values given as a prediction. The range is chosen so that statistically such predictions will be correct a certain percentage of the time, typically 95% of the time. |
| **Digital Divide** | A difference between two groups within or between nations in terms of resources, infrastructure, and access to digital information and communications technology. |
| **Disaggregating Data** | When disaggregating data, you separate the data based on one variable in order to compare values of another variable between groups. |
| **Edge or Link** | Connection between two nodes in a graph. |
| **Graph** | A network of nodes connected by links. |
| **Inferential Statistics** | Makes predictions about a population of measurements based on a sample of measurements. |
| **Latitude** | Position on the Earth north-to-south, measured in degrees from -90 (south pole) to +90 (north pole). |
| **Longitude** | Position on the Earth East/West measured in degrees from -180 to +180, with 0 passing through England. |
| **Model** | Mathematical constants, relationships, and algorithms used to describe something real. A model includes some patterns of the real thing and abstracts away other details. |
| **Monte Carlo Simulation** | Using a random number generator to observe the range of results that occur in a simulation that has random variability. |
| **Narrow AI** | Artificial intelligence in which a program is written to accomplish a specific task. |
| **Node** | A point in a graph, the kind of graph that describes a network. |
| **Parameter** | A value used in a model. |
| **p-value** | The p-value, or just p, is the probability that an apparent pattern (such as the value of variable being associated with the value of another variable) has arisen by chance and not due to a pattern that actually exists. Typically p<0.05 is accepted as scientific evidence. |
| **Simulation** | Use of a model to generate meaningful but fictional data. |
| **Strong AI** | Artificial intelligence with a generalized learning ability, on par with human intelligence. |
| **Training Set** | A data set in which the input and the desired output are both provided to the computer. |
| **Validation Set** | A data set in which the input is provided to the computer and the desired output is known, so that it can be determined how well a machine learning an algorithm is working. |