

accuracy - Accuracy is an important factor in assessing the success of data mining. When applied to data, accuracy refers to the rate of correct values in the data. When applied to models, accuracy refers to the degree of fit between the model and the data. This measures how error-free the model's predictions are. Also see precision.

activation function - A function used by a node in a neural net to transform input data from any domain of values into a finite range of values. The original idea was to approximate the way neurons fired, and the activation function took on the value 0 until the input became large and the value jumped to 1. The discontinuity of this 0-or-1 function caused mathematical problems, and sigmoid-shaped functions (e.g., the logistic function) are now used.

agglomerative clustering -An unsupervised technique where each data instance initially represents its own cluster. Successive iterations of the algorithm merge pairs of highly similar clusters until all instances become members of a single cluster. In the last step, a decision is made about which clustering is a best final result.

ANN - Artificial Neural Network (see neural network).

antecedent -When an association between two variables is defined, the first item (or left-hand side) is called the antecedent. For example, in the relationship "When a consumer buys milk, they buy bread 14% of the time," "buys milk" is the antecedent. (See consequent).

API -An application program interface. When a software system features an API, it provides a means by which programs written outside of the system can interface with the system to perform additional functions. For example, a data mining software system may have an API which permits user-written programs to perform such tasks as extract data, perform additional statistical analysis, create specialized charts, generate a model, or make a prediction from a model.

Apriori Rule Algorithm -In association rule development, one of two major algorithms (General Rule Induction is a second algorithm). The Apriori algorithm analyses a data set to determine which combination of items occurs frequently. This is an iterative algorithm that uses prior knowledge as an indicator of future events. (See also confidence, support and improvement)

Apriori Segmentation - A segmentation methodology that groups data based upon some factor or factors that are known or believed in advance.

association rule - A production rule whose consequent may contain multiple conditions and attribute relationships.

associations -An association algorithm creates rules that describe how often events have occurred together. For example, "When consumers buy milk, they also buy bread 14% of the time." There are two major association algorithms _ Apriori and General Rule Induction (GRI). This approach is often referenced as Market Basket Analysis (MBA).

average member technique -An unsupervised clustering neural network explanation where the most typical member of each cluster is computed by finding the average value for each class attribute.

back-propagation -A training method used to calculate the weights in a neural net from the data. It is one of the most common methods used to train neural networks. The key aspect of back-propagation is that hidden nodes are able to determine how to change weights if they receive error information from each of the output nodes. Through this information flow, neural networks adjust their weights in order to produce the most efficient model (lowest error). (see Gradient Descent)

bagging - A supervised learning approach that allows several models to have an equal vote in the classification of new instances.

Bayes theorem - The probability of a hypothesis given some evidence is equal to the probability of the evidence given the hypothesis, times the probability of the hypothesis, divided by the probability of the evidence.

Bayesian Information Criteria (BIC) - The BIC gives the posterior odds for one data mining model against another model assuming neither model is favored initially.

bias -In a neural network, bias refers to the constant terms in the model. (Note that bias has a different meaning to most data analysts.) Also see precision.

binning -A data preparation activity that converts continuous data to discrete data by replacing a value from a continuous range with a bin identifier, where each bin represents a range of values. For example, age could be converted to bins such as 20 or under, 21-40, 41-65 and over 65.

boosting -A supervised learning approach that allows several models to take part in the classification of new instances. Each model has an associated weight that is applied toward new instance classification.

bootstrapping -Training data sets are created by re-sampling with replacement from the original training set, so data records may occur more than once. In other words, this method treats a sample as if it were the entire population. Usually, final estimates are obtained by taking the average of the estimates from each of the bootstrap test sets.

Brute Force Algorithms - A computer technique that exhaustively uses the repetition of very simple steps repeated in order to find an optimal solution. They stand in contrast to complex techniques that are less wasteful in moving toward an optimal solution but are harder to construct and are more computationally expensive to execute.

Cardinality -The number of different values a categorical predictor or OLAP dimension can have. High cardinality predictors and dimensions have large numbers of different values (e.g., zip codes), low cardinality fields have few different values (e.g., eye color).

CART (Classification and Regression Trees) -CART is a method of splitting the independent variables into small groups and fitting a constant function to the small data sets. Predictors are picked as they decrease the disorder in the data. In building a CART model, each predictor is picked based on how well it teases apart the records with different predictions.

categorical data -Categorical data fits into a small number of discrete categories (as opposed to continuous). Categorical data is either non-ordered (nominal) such as gender or city, or ordered (ordinal) such as high, medium, or low temperatures.

CHAID (Chi Square Automatic Interaction Detection) -An algorithm for fitting categorical trees. It relies on the chi-squared statistic to split the data into small connected data sets.

chi-square -A statistic that assesses how well a model fits the data by comparing actual to expected distributions. In data mining, it is most commonly used to find homogeneous subsets for fitting categorical trees as in CHAID.

classification -Refers to the data mining problem of attempting to predict the category of categorical data by building a model based on some predictor variables.

classification tree - A decision tree that places categorical variables into classes.

cleaning (cleansing) -Refers to a step in preparing data for a data mining activity. Obvious data errors are detected and corrected (e.g., improbable dates) and missing data is replaced.

clustering -Clustering algorithms find groups of items that are similar. For example, clustering could be used by an insurance company to group customers according to income, age, types of policies purchased and prior claims experience. It divides a data set so that records with similar content are in the same group, and groups are as different as possible from each other. Since the categories are unspecified, this is sometimes referred to as unsupervised learning or indirect data mining.

Conditional Probability $p(x|y)$ -The probability of an event happening given that some event has already occurred. For example, the probability of a person buying a Certificate of Deposit is greater if the person has previously purchased a Certificate of Deposit.

confidence -Within data mining, confidence has a different meaning than traditional statistics. In the former, it is used when discussing association rules. Confidence of rule "B given A" is a measure of how much more likely it is that B occurs when A has occurred. It is expressed as a percentage, with 100% meaning B always occurs if A has occurred. Statisticians refer to this as the conditional (posterior) probability of B given A. When used with association rules, the term confidence is observational rather than predictive. (Statisticians also use this term in an unrelated way. There are ways to estimate an interval and the probability that the interval contains the true value of a parameter is called the interval confidence. So a 95% confidence interval for the mean has a probability of .95 of covering the true value of the mean.)

confusion matrix -A confusion matrix shows the counts of the actual versus predicted class values. It shows not only how well the model predicts, but also presents the details needed to see exactly where things may have gone wrong.

concept hierarchy - A mapping that allows attributes to be viewed from varying levels of detail.

consequent -When an association between two variables is defined, the second item (or right-hand side) is called the consequent. For example, in the relationship "When a consumer buys bread, he buys a milk 14% of the time," "buys milk" is the consequent.

constellation schema - A variation of the star schema that allows more than one central fact table.

continuous -Continuous data can have any value in an interval of real numbers. That is, the value does not have to be an integer. Continuous is the opposite of discrete or categorical.

coverage - The number or percentage of times that a rule can be applied.

cross validation -A method of estimating the accuracy of a classification or regression model. The data set is divided into several parts, with each part in turn used to test a model fitted to the remaining parts.

Customer Relationship Management (CRM) - A business strategy for managing customer experiences with a company. It involves the integration of sales management, analytics, segmentation and technology in order to provide customers with a seamless view of the corporation. For a corporation,

it provides a unique view of how customers use your services, customer value and market efficiency.

data -Values collected through record keeping or by polling, observing, or measuring, typically organized for analysis or decision making. More simply, data is facts, transactions and figures.

data format -Data items can exist in many formats such as text, integer and floating-point decimal. Data format refers to the form of the data in the database.

data mining -An information extraction activity whose goal is to discover hidden facts contained in databases. Using a combination of machine learning, statistical analysis, modeling techniques and database technology, data mining finds patterns and subtle relationships in data and infers rules that allow the prediction of future results. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis.

data mining method - Procedures and algorithms designed to analyze the data in databases.

data model - A notational language that documents the structure of data independent of how the data will be used.

data warehouse - A historical database designed for decision support.

DBMS - Database management systems.

decision tree - A tree-like way of representing a collection of hierarchical rules that lead to a class or value.

deduction - Deduction infers information that is a logical consequence of the data.

degree of fit - A measure of how closely the model fits the training data. A common measure is r-square.

dependent variable -The dependent variables (outputs or responses) of a model are the variables predicted by the equation or rules of the model using the independent variables (inputs or predictors).

deployment -After the model is trained and validated, it is used to analyze new data and make predictions. This use of the model is called deployment.

dimension -Each attribute of a case or occurrence in the data being mined. Stored as a field in a flat file record or a column of relational database table.

dimensional table - A relational table containing information about one of the dimensions of a star schema.

discrete - A data item that has a finite set of values. Discrete is the opposite of continuous.

discriminant analysis - A statistical method based on maximum likelihood for determining boundaries that separate the data into categories.

embedded data mining -An implementation of data mining where the algorithms are embedded into existing data stores and information delivery processes rather than requiring data extraction and new data stores.

entity relationship design (ERD) -A data modeling tool that shows the structure of the data in terms of the entities and relationships between entities. Relationships between entities can be one-to-one, one-to-many, or many-to-many.

entity - A generic representation for a class of persons, places or things.

entropy -A way to measure variability other than the variance statistic. Some decision trees split the data into groups based on minimum entropy.

epoch - One complete pass of the training data through a neural network.

error rate - A number that reflects the rate of errors made by a predictive model. It is one minus accuracy.

expert system - A data processing system comprising a knowledge base (rules), inference (rules) and a working memory.

exploratory analysis -Looking at data to discover relationships not previously detected. Exploratory analysis tools typically assist the user in creating tables and graphical displays.

external data -Data not collected by the organization, such as data available from a reference book, a government source or a proprietary database.

fact table - A relational table that defines the dimensions of the multidimensional space with a star schema.

factor analysis - A statistical technique which seeks to reduce the number of total predictions from a large number to only a few "factors" that have the majority of impact on the predicted outcome.

field -The structural component of a database that is common to all records. All fields have values. Other names are attributes, variables, features, dimensions, or table columns.

first normal form - A rule that requires all attributes within an entity to have a single value.

feed-forward - A neural net in which the signals only flow in one direction, from the inputs to the outputs.

fuzzy logic - Fuzzy logic is applied to fuzzy sets where membership in a fuzzy set is a probability, not necessarily 0 or 1. Non-fuzzy logic manipulates outcomes that are either true or false. Fuzzy logic needs to be able to manipulate degrees of "maybe" in addition to true and false.

genetic algorithms - A computer-based method of generating and testing combinations of possible input parameters to find the optimal output. It uses processes based on natural evolution concepts such as genetic combination, mutation and natural selection. This algorithm was initially developed by John Holland (1986).

Gini Metric -A measure of the disorder reduction caused by the splitting of data in a decision tree algorithm. Gini and the entropy metric are the most popular ways of selecting predictors in the CART decision tree algorithm.

granularity - A term used to describe the level of detail of stored information.

GUI - Graphical User Interface.

hidden nodes -The nodes in the hidden layers in a neural net. Unlike input and output nodes, the number of hidden nodes is not predetermined. The accuracy of the resulting model is affected by the number of hidden nodes. Since the number of hidden nodes directly affects the number of parameters in the model, a neural net needs a sufficient number of hidden nodes to enable it to properly model the underlying behavior. On the other hand, a net with too many hidden nodes will over fit the data. Some neural net products include algorithms that search over a number of alternative neural nets by varying the number of hidden nodes, in the end choosing the model that gets the best results without over fitting.

hill climbing search - A simple optimization technique that modifies a proposed solution by a small amount and then accepts it if it is better than the prior solution. This approach is slow and suffers from being caught in local optima.

independent variable -The independent variables (inputs or predictors) of a model are the variables used in the equation or rules of the model to predict the output (dependent) variable.

induction - A technique that infers generalizations from the information in the data.

interaction - Two independent variables interact when changes in the value of one change the effect on the dependent variable of the other.

internal data - Data collected by an organization such as operating and customer data.

intelligent agent -A software application that assists a system or user by automating a task. Intelligent agents must recognize events and use domain knowledge to take appropriate actions based on those events.

k-nearest neighbor -A classification method that classifies a point by calculating the distances between the point and points in the training data set. Then it assigns the point to the class that is most common among its k-nearest neighbors (where k is an integer).

Kohonen feature map -A type of neural network that uses unsupervised learning to find patterns in data. In data mining it is employed for cluster analysis. Teuvo Kohonen (1982) first formalized unsupervised clustering in the 1980s when he introduced Kohonen feature maps.

layer -Nodes in a neural net are usually grouped into layers, with each layer described as input, output or hidden. There are as many input nodes as there are input (independent) variables and as many output nodes as there are output (dependent) variables. Typically, there are one or two hidden layers.

leaf - A node not further split -- the terminal grouping -- in a classification or decision tree.

learning - Training models (estimating their parameters) based on existing data.

least squares -The most common method of training (estimating) the weights (parameters) of a model by choosing the weights that minimize the sum of the squared deviation of the predicted values of the model from the observed values of the data.

left-hand side -When an association between two variables is defined, the first item is called the left-hand side (or antecedent). For example, in the relationship "When a consumer buys milk, they buy

bread 14% of the time", "buys a milk" is the left-hand side.

lift -The probability of class C_i given a sample taken from population P divided by the probability of C_i given the entire population P .

logistic regression (logistic discriminant analysis) -A generalization of linear regression. It is used for predicting a binary variable (with values such as yes/no or 0/1). An example of its use is modeling the odds that a borrower will default on a loan based on the borrower's income, debt and age.

machine learning - A field of science and technology concerned with building machines that learn. In general, it differs from Artificial Intelligence in that learning is considered to be just one of a number of ways of creating artificial intelligence.

many-to-many relationship -A relationship between two entities, A and B , where each instance of A is associated with one or several instances of B , and each instances of B is associated one or several instances of A .

market basket analysis (MBA) -A data mining strategy used for finding groups of items that tend to occur together in a transaction (or market basket). This technique provides the likelihood of different products being purchased together and is expressed as rules.

MARS - Multivariate Adaptive Regression Splines. MARS is a generalization of a decision tree.

maximum likelihood - Another training or estimation method. The maximum likelihood estimate of a parameter is the value of a parameter that maximizes the probability that the data came from the population defined by the parameter.

mean - The arithmetic average value of a collection of numeric data.

median -The value in the middle of a collection of ordered data. In other words, the value with the same number of items above and below it.

memory based reasoning (MBR) -A technique for comparing records in a database by comparing them with similar records that are already classified. A form of nearest neighbor.

minimum description length (MDL) principal -The idea that the least complex predictive model (with acceptable accuracy) will be the one that best reflects the true underlying model and performs most accurately on the data.

missing data -Data values can be missing because they were not measured, not answered, were unknown or were lost. Data mining methods vary in the way they treat missing values. Typically, they ignore the missing values, or omit any records containing missing values, or replace missing values with the mode or mean, or infer missing values from existing values.

mode - The most common value in a data set. If more than one value occurs the same number of times, the data is multi-modal.

model -An important function of data mining is the production of a model. A model can be descriptive or predictive. A descriptive model helps in understanding underlying processes or behavior. For example, an association model describes consumer behavior. A predictive model is an equation or set of rules that makes it possible to predict an unseen or unmeasured value (the dependent variable or output) from other, known values (independent variables or input). The form

of the equation or rules is suggested by mining data collected from the process under study. Some training or estimation technique is used to estimate the parameters of the equation or rules.

MPP -Massively parallel processing, a computer configuration that is able to use hundreds or thousands of CPUs simultaneously. In MPP each node may be a single CPU or a collection of SMP CPUs. An MPP collection of SMP nodes is sometimes called an SMP cluster. Each node has its own copy of the operating system, memory, and disk storage, and there is a data or process exchange mechanism so that each computer can work on a different part of a problem. Software must be written specifically to take advantage of this architecture.

nearest neighbor - A data mining technique that performs prediction by finding the prediction value of records (near neighbors) similar to the record to be predicted.

neural network -A complex nonlinear modeling technique based on a model of a human neuron. A neural net is used to predict outputs (dependent variables) from a set of inputs (independent variables) by taking linear combinations of the inputs and then making nonlinear transformations of the linear combinations using an activation function. It can be shown theoretically that such combinations and transformations can approximate virtually any type of response function. Thus, neural nets use large numbers of parameters to approximate any model. Neural nets are often applied to predict future outcome based on prior experience. For example, a neural net application could be used to predict who will respond to a direct mailing.

node -A decision point in a classification (i.e., decision) tree. Also, a point in a neural net that combines input from other nodes and produces an output through application of an activation function.

nominal categorical predictor -A predictor that is categorical (finite cardinality) but where the values of the predictor have no particular order. For example, eye color.

noise -The difference between a model and its predictions. Sometimes data is referred to as noisy when it contains errors such as many missing or incorrect values or when there are extraneous columns.

non-applicable data - Missing values that would be logically impossible (e.g., pregnant males) or are obviously not relevant.

normalize -A collection of numeric data is normalized by subtracting the minimum value from all values and dividing by the range of the data. This yields data with a similarly shaped histogram but with all values between 0 and 1. It is useful to do this for all inputs into neural nets and also for inputs into other regression models. (Also see standardize.)

Occam's Razor - A rule of thumb used by many scientists that advocates favoring the simplest that adequately explains (or predicts) an event. This is more formally captured for machine learning and data mining as the minimum description length principle.

OLAP - On-Line Analytical Processing tools give the user the capability to perform multi-dimensional analysis of the data.

one-to-one relationship - A relationship between two entities, A and B, where each instance is associated with exactly one instance of B.

one-to-many relationship - A relationship between two entities, A and B, where each instance of A is

associated with one or several instances of B.

OLTP - On-line transaction processing are database procedures designed to process individual transactions.

operational database - A database designed for processing the day-to-day transactions of a company.

optimization criterion -A positive function of the difference between predictions and data estimates that are chosen so as to optimize the function or criterion. Least squares and maximum likelihood are examples.

ordinal categorical predictors -A categorical predictor (i.e., has a finite number of values) where the values have order but do not convey meaningful intervals or distances between them. For example, high, medium and low income.

outliers -Technically, outliers are data items that did not (or are thought not to have) come from the assumed population of data -- for example, a non-numeric when you are expecting only numeric values. A more casual usage refers to data items that fall outside the boundaries that enclose most other data items in the data set.

over fitting -A tendency of some modeling techniques to assign importance to random variations in the data by declaring them important patterns.

overlay -Data not collected by the organization, such as data from a proprietary database, that is combined with the organization's own data.

parallel processing - Several computers or CPUs linked together so that each can be computing simultaneously.

pattern - Analysts and statisticians spend much of their time looking for patterns in data. A pattern can be a relationship between two variables. Data mining techniques include automatic pattern discovery that makes it possible to detect complicated non-linear relationships in data. Patterns are not the same as causality.

precision -The precision of an estimate of a parameter in a model is a measure of how variable the estimate would be over other similar data sets. A very precise estimate would be one that did not vary much over different data sets. Precision does not measure accuracy. Accuracy is a measure of how close the estimate is to the real value of the parameter. Accuracy is measured by the average distance over different data sets of the estimate from the real value. Estimates can be accurate but not precise, or precise but not accurate. A precise but inaccurate estimate is usually biased, with the bias equal to the average distance from the real value of the parameter.

predictability - Some data mining vendors use predictability of associations or sequences to mean the same as confidence.

predictor - A column or field in a database that could be used to build a model to predict the values in another field or column.

prevalence -The measure of how often the collection of items in an association occur together as a percentage of all the transactions. For example, "In 2% of the purchases at the hardware store, both a pick and a shovel were bought."

principal component analysis -A data analysis technique that seeks to weight the importance of a variety of predictors so that they optimally discriminate between various possible predicted outcomes.

prior probability -The probability of an event occurring without dependence on (conditional to) some other event. In contrast to conditional probability.

pruning -Eliminating lower level splits or entire sub-trees in a decision tree. This term is also used to describe algorithms that adjust the topology of a neural net by removing (i.e., pruning) hidden nodes.

radial basis functions -Neural networks that combine some of the advantages of neural networks with those of nearest neighbor techniques. In radial basis functions the hidden layer is made up of nodes that represent prototypes or clusters of records.

range -The range of the data is the difference between the maximum value and the minimum value. Alternatively, range can include the minimum and maximum, as in "The value ranges from 2 to 8."

record -The fundamental data structure used for performing data analysis. Also called a table row or example. A typical record would be the structure that contains all relevant information pertinent to one particular customer or account.

RDBMS - Relational Database Management System.

regression -A data analysis technique classically used in statistics for building predictive models for continuous prediction fields. The technique automatically determines a mathematical equation that minimizes some measure of the error between the prediction from the regression model and the actual data.

regression tree - A decision tree that predicts values of continuous variables.

reinforcement learning -A training model where an intelligence engine (e.g. neural network) is presented with a sequence of input data followed by a reinforcement signal.

relational database - A database built to conform to the relational data model; includes the catalog and all the data described therein.

response - A binary prediction field that indicates response or non response to a variety of marketing interventions. The term is generally used when referring to models that predict response or to the response field itself.

resubstitution error - The estimate of error based on the differences between the predicted values of a trained model and the observed values in the training set.

right-hand side -When an association between two variables is defined, the second item is called the right-hand side (or consequent). For example, in the relationship "When a consumer buys milk, he buys bread 14% of the time," "buys bread" is the right-hand side.

r-squared -A number between 0 and 1 that measures how well a model fits its training data. One is a perfect fit; however, zero implies the model has no predictive ability. It is computed as the covariance between the predicted and observed values divided by the standard deviations of the predicted and observed values.

sampling - Creating a subset of data from the whole. Random sampling attempts to represent the whole by choosing the sample through a random mechanism. Sampling can provide relatively good models at much less computational expense than using the entire database.

second normal form (2NF) - A entity is in 2NF if it is in 1NF and all nonkey attributes are dependent on the full primary key.

segmentation -The process or result of the process that creates mutually exclusive collections of records that share similar attributes either in unsupervised learning (such as clustering) or in supervised learning for a particular prediction field.

sensitivity analysis -The process which determines the sensitivity of a predictive model to small fluctuations in predictor value. Through this technique end users can gauge the effects of noise and environmental change on the accuracy of the model. Varying the parameters of a model to assess the change in its output.

sequence discovery - The same as association, except that the time sequence of events is also considered. For example, "Twenty percent of the people who buy a VCR buy a camcorder within four months."

sigmoid function -One of several commonly used neural network evaluation functions. The sigmoid function is continuous and outputs values between 0 and 1.

significance -A probability measure of how strongly the data support a certain result (usually of a statistical test). If the significance of a result is said to be .05, it means that there is only a .05 probability that the result could have happened by chance alone. Very low significance (less than .05) is usually taken as evidence that the data mining model should be accepted since events with very low probability seldom occur. So if the estimate of a parameter in a model showed a significance of .01 that would be evidence that the parameter must be in the model.

SMP -Symmetric multi-processing is a computer configuration where many CPUs share a common operating system, main memory and disks. They can work on different parts of a problem at the same time.

standardize -A collection of numeric data is standardized by subtracting a measure of central location (such as the mean or median) and by dividing by some measure of spread (such as the standard deviation, interquartile range or range). This yields data with a similarly shaped histogram with values centered around 0. It is sometimes useful to do this with inputs into neural nets and also inputs into other regression models. (Also see normalize.)

snowflake schema -A variation of the star schema where some of the dimension tables linked to the fact table are further subdivided. This permits the dimension tables to be normalized, which means less total storage.

star schema -A multidimensional data warehouse model implemented within a relational database. The model consists of a fact table and one or more dimension tables.

structured query language (SQL) - A standard language for accessing data in relational databases.

supervised learning -The collection of techniques where analysis uses a well-defined (known) dependent variable. All regression and classification techniques are supervised.

support - The measure of how often the collection of items in an association occur together as a percentage of all the transactions. For example, "In 2% of the purchases at the grocery store, both bread and milk were bought."

test data - A data set independent of the training data set, used to fine-tune the estimates of the model parameters (i.e., weights).

test error - The estimate of error based on the difference between the predictions of a model on a test data set and the observed values in the test data set when the test data set was not used to train the model.

third normal form (3NF) - A entity is in 3NF if it is in 2NF and every nonkey attribute is dependent entirely of the primary key.

time series - A series of measurements taken at consecutive points in time. Data mining products which handle time series incorporate time-related operators such as moving average. (Also see windowing.)

time series model - A model that forecasts future values of a time series based on past values. The model form and training of the model usually take into consideration the correlation between values as a function of their separation in time.

topology - For a neural net, topology refers to the number of layers and the number of nodes in each layer.

training - Another term for estimating a model's parameters based on the data set at hand.

training data - A data set used to estimate or train a model.

transformation - A re-expression of the data such as aggregating it, normalizing it, changing its unit of measure, or taking the logarithm of each data item.

unsupervised learning - This term refers to the collection of techniques where groupings of the data are defined without the use of a dependent variable. Cluster analysis is an example.

validation - The process of testing the models with a data set different from the training data set.

variance - The most commonly used statistical measure of dispersion. The first step is to square the deviations of a data item from its average value. Then the average of the squared deviations is calculated to obtain an overall measure of variability.

visualization - Visualization tools graphically display data to facilitate better understanding of its meaning. Graphical capabilities range from simple scatter plots to complex multi-dimensional representations.

windowing - Used when training a model with time series data. A window is the period of time used for each training case. For example, if we have weekly stock price data that covers fifty weeks, and we set the window to five weeks, then the first training case uses weeks one through five and compares its prediction to week six. The second case uses weeks two through six to predict week seven, and so on.