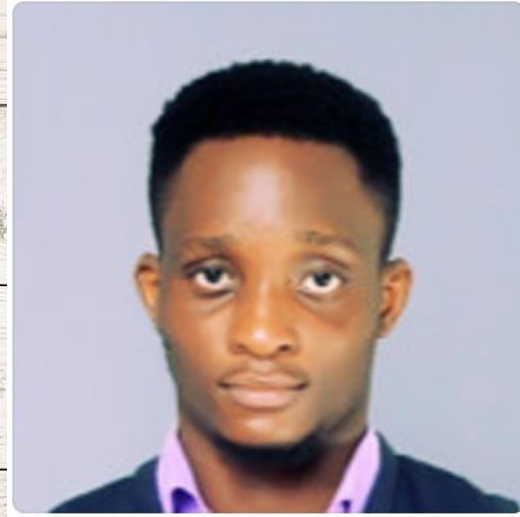# Machine Learning Glossary
#26DaysOfConcepts

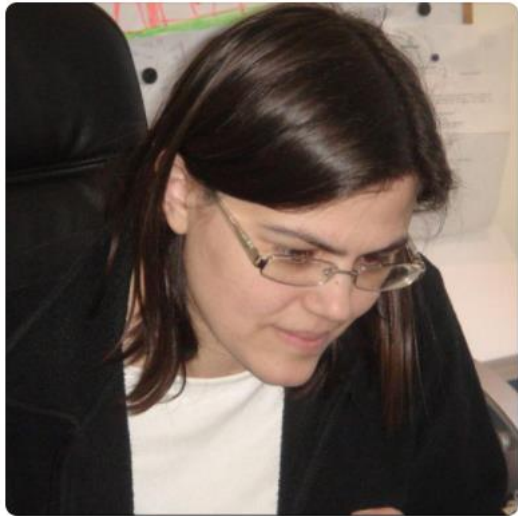## Udacity-Microsoft Azure Machine Learning Scholarship 2020
#sg_spaic

Prepared By

Memphis Ndego

Verrah Otiende

Dimitra Karamperi

Sadmi Bouhafs

Accuracy is a metric for the evaluation of classification models and is calculated as the number of correctly classified items divided by the total number of items in the test set. Ranges from 0 (least accurate) to 1 (most accurate).

Accountability as per Microsoft AI principles, algorithms and the people who write them should be responsible or answerable for their impacts. Accountability is a foundational principle, along with transparency.

AlexNet is the name of a CNN, designed by Alex Krizhevsky. It competed in the ImageNet Large Scale Visual Recognition Challenge on September 30, 2012. The network achieved a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up.

Algorithm is a step by step procedure that frequently involves repetition of an operation to solve logical and mathematical problems. The words 'algorithm' and 'algorism' come from the name of a Persian mathematician called Al-Khwarizmi.

Anomaly Detection is a machine learning technique that identifies events or observations that vary significantly from the majority of the data. Essentially, it assumes that the majority of entities are normal; anything that differs is an anomaly.

AutoML (Automated Machine Learning) is an operation that automates the entire machine learning process. It intelligently tests multiple algorithms and hyperparameters in parallel and returns the best one.

Bagging helps reduce overfitting for models that tend to have high variance and uses random subsampling of the training data to produce a bag of trained models. The resulting trained models are homogeneous, while the final prediction is an average prediction from individual models.

Batch inferencing is the process of generating predictions on a batch of observations. The batch jobs are typically generated on some recurring schedule. These predictions are then stored in a database and can be made available to developers or end users.

Bias measures how inaccurate the model prediction is in comparison with the true output. It is an error that results from erroneous assumptions made to simplify the training process. High model complexity tends to have a low bias.

Binning is a feature-engineering task. It is the process of grouping individual data values into a smaller number of "bins" or groups according to defined criteria. The most frequent example is binning values for age into categories like 20-39, 40-59, and 60-79.

Bokeh is an interactive visualization library for modern web browsers. It provides quick, elegant, and interactive plots, dashboards, and data applications with high-performance interactivity over very large or streaming datasets.

Boosting is a machine learning technique that tries to reduce the bias. In this process, trained machine learning models(weak learners) combine their outputs to produce another model(strong learner). Learning is done sequentially: each new learner corrects errors from their previous learners.

Bayes Rule also known as Bayes's

theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. For example, if the risk of developing health problems is known to increase with age, Bayes's theorem allows the risk to an individual of a known age to be assessed more accurately than simply assuming that the individual is typical of the population as a whole. The formula is:

$$P(A/B) = \frac{P(B/A)*P(A)}{P(B)}$$

Classification is a supervised machine learning approach that categorizes a set of data into classes. It can be performed on both structured and unstructured data. The output from a classification problem is categorical: binary (two categories) or multi-class (more than two categories).

Clustering is an unsupervised machine learning approach that segregates data points into groups (clusters) such that data points in the same group are more similar compared to data points in other groups. Essentially, the goal of a clustering problem is to maximize both intra-cluster similarity and inter-cluster differences.

Compute instance - Azure is a fully-managed cloud-based workstation that gives you access to various development environments, such as Jupyter Notebooks where you start writing your Data Science Code.

Compute target - Azure is a designated compute resource/environment where you run your training script or host your service deployment. This location may be your local machine or a cloud-based compute resource. Using compute targets makes it easy for you to later change your compute environment without having to change your code.

Confusion matrix is a table that allows visualization of the performance of a classification model on a set of test data for which the true values are known. It has two rows and two columns with the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN).

Cost function measures the performance of an ML model for given data. It is a numerical representation of the error between the predicted value and the actual value. The most commonly used cost function for linear regression is the root mean squared error (RMSE).

Chatbot is a software application used to conduct an on-line chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent. Designed to convincingly simulate the way a human would behave as a conversational partner, chatbot systems typically require continuous tuning and testing, and many in production remain unable to adequately converse or pass the industry standard Turing test.

Covariance is a statistical term that indicates the direction of the linear relationship between two variables. By direction we mean if the variables are directly proportional or inversely proportional to each other. (Increasing the value of one variable might have a positive or a negative impact on the value of the other variable). In probability, covariance is the measure of the joint probability for two random variables.

Correlation is a statistical term that is used to study the strength of a relationship between two, numerically measured, continuous variables. The correlation coefficient is a dimensionless metric and its value ranges from -1 to +1. The closer it is to +1 or -1, the more closely the two variables are related. If there is no linear relationship at all between two variables, then the correlation coefficient will certainly be 0.

Data augmentation is the process of using computer algorithms or other synthetic means to increase the size of a collected dataset. As ML algorithms become more resistant to overfitting when they are trained with more data, it is possible to significantly increase the size of a dataset by transforming existing data to create additional data.

Datasets are resources for exploring, transforming, and managing data in Azure ML. A dataset is essentially a reference that points to the data in storage. It is used to get specific data files in the datastores. One point to always remember is that they are not copies of the data but references that point to the original data.

Datastore is a layer of abstraction over the supported Azure storage services. It stores all the information needed to connect to a particular storage service and it provides the answer to the very important question of how one can connect in a secure way with the storage that keeps their data. It is an increased level of security.

Discrete features are discontinuous and can take only definite values from a set of possible values. In ML discrete features are commonly in the form of nominal – with no numerical relationships between labels, ordinal – ranked or sorted, and binary – can only be one of two labels.

Data drift occurs when there is a change in input data for a model. Overtime, this change results in degradation in the model's performance. Common causes include changes in upstream process, natural drift, data quality, and change in relationship between features. Dataset monitors are set up to detect data drift.

Data poisoning is a conflicting attack that manipulates the training dataset with the intention of controlling prediction behavior of the trained model in a way that the boundary labels shift into the desired class and not the required class.

Decision trees are non-parametric supervised learning algorithms used for both classification and regression tasks. The main idea is to find descriptive features which contain the most information regarding the target feature. Decision tree is made up of a root node, interior nodes, and leaf nodes connected by branches.

Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness. While the act of faking content is not new, deepfakes leverage powerful techniques from machine learning and artificial intelligence to manipulate or generate visual and audio content with a high potential to deceive.

Deployment is an ML process of integrating the model to an existing production environment so that it takes in an input to return an output. It makes predictions from a trained ML model available to other users.

Dimensionality reduction is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension.

Decision boundary is a hypersurface that partitions the underlying vector space into two sets, one for each class. The classifier will classify all the points on one side of the decision boundary as belonging to one class and all those on the other side as belonging to the other class.

Data is any unprocessed fact, value, text, sound or picture not interpreted or analyzed. It is the most important part of ML and without it training and automation of models is not possible.

Deep learning is a subset of ML where ANN algorithms inspired by the human brain learn to perform classification tasks from large amounts of data. It is referred to as DL because the neural networks have deep layers that enable learning and can achieve state-of-the-art accuracy sometimes exceeding human-level performance. DL is the key technology behind driverless cars, voice control in consumer devices.

Designer - Azure Machine Learning is a visual-first environment that has modules for data transformation, model training and evaluation and lets you create and publish ML pipelines via a drag and drop interface without needing to write a single line of code.

Early stopping is a regularization procedure used for prevention of overfitting in an iterative process such as gradient descent when training a learner. The procedure updates the learner in each iteration to make a better fit in terms of performance and desirable accuracy.

Embedding is a low-dimensional space for translating high-dimensional vectors like sparse vectors representing words. The process is a learned continuous vector representation of discrete variables which makes ML models efficient and easier to reuse across models.

Endpoints - Azure are a cloud service for machine learning that exposes real-time endpoints for scoring as well as pipelines for advanced automation.

Encoding machine learning algorithms need to have data in numerical form. Thus, when we have categorical data, we need to encode it in some way so that it is represented numerically. There are two common approaches for encoding categorical data: ordinal encoding and one hot encoding.

Epoch is the number of times the algorithm sees the entire data set. The number of epochs is usually large (hundreds or thousands) allowing the learning algorithm to run until the model error is minimized. However, there is no magic rule for choosing the number of epochs. It is a hyperparameter that must be determined before training begins.

Evaluation is an integral component of any data science project. Model evaluation aims to estimate the generalization accuracy of a model on unseen or out-of-sample data. Common evaluation metrics include accuracy, precision, recall, F1-score, ROC and others, while their use depends heavily on the task at hand.

Ensemble learning is a powerful machine learning technique that combines the decisions from multiple models to produce a predictive model with improved performance. It helps to reduce model bias, variance, and noise. The most popular ensemble methods include bagging, boosting, and stacking.

Environments - Azure are software applications that provide a whole suite of tools designed to help a developer build out projects. Common environments include Jupyter Notebook, Azure Notebook, Azure Databricks, Visual Studio Code, and Visual Studio.

F₁-score is a measure of a test's accuracy in binary classification. It is calculated from the precision and recall of the test, the formula is:

$$2*\frac{(Precision+Recall)}{(Precision*Recall)}.$$

Fairness is one of the core responsible AI principles. In machine learning, a given algorithm is said to be **fair**, or to have **fairness**, if its results are independent of given variables, especially those considered sensitive, such as the traits of individuals which should not correlate with the outcome (i.e. gender, ethnicity, disability, etc).

False positive is an error specific to classification problems. It results when a model incorrectly predicts a negative class as positive. It is also known as type I error.

False negative is an error specific to classification problems. It results when a model incorrectly predicts a positive class as negative. It is also known as type II error.

Forecasting involves making predictions based on past and present analysis of trends to derive actionable insights. A commonplace example might be self-learning algorithms to analyze the past impact of currency fluctuations then envisage better real-time forecasts.

Feature is an input variable used in making predictions.

Frameworks are interfaces, libraries or developer tools for building ML models with minimal complexity of underlying algorithms. The most common frameworks for ML are TensorFlow, PyTorch, Scikit-Learn, Azure ML studio, AWS ML.

Feature engineering is the process of manipulating existing features to create new features, with the goal of improving model training. It can be as simple as applying a mathematical function to a feature (e.g. adding 1 to all values in an existing feature) or it can be as complex as training a separate machine learning model to create values for new features.

Flagging is a feature engineering task that amounts to deriving boolean conditions that are expressed through boolean values (0/1 or True/False) for each entity.

Gain and Lift charts are used to evaluate the performance of different classification models. They help to find the best predictive model among multiple challenger models. The higher the lift (the further up it is from the baseline), the better the model. Gain > 1 means the results from the predictive model are better than the random model.

Generative Adversarial Networks(GANs) are generative models: they create new data instances that resemble your training data. Two deep networks are trained simultaneously by an adversarial process. A generator (the artist) learns to create images that look real, while a discriminator (the art critic) learns to tell real images from fake.

Gated Recurrent Unit(GRU) is a gating mechanism that aims to solve the vanishing gradient problem often encountered in Recurrent Neural Network. GRUs are able to solve this problem by using an update gate that controls the information that flows into memory and a reset gate that controls the information that flows out of memory.

Gaussian distribution is a probability distribution which accurately models a large number of phenomena in the world. In machine learning, it may be used to find non-linear regressions as well as to reduce dimensionality by identifying which dimensions of a dataset have larger variance and thus may contain more useful information.

Generalization curve is a loss curve showing both the training set and the validation set. A generalization curve can help detect possible overfitting, for example, when loss for the validation set is significantly higher than for the training set.

Generator is the subsystem within a Generative Adversarial Network that learns to create fake data by incorporating feedback from the other subsystem of GAN, the discriminator. It learns to make the discriminator classify its output as real.

GPUs are programmable processors designed to drive high-quality graphics processing. They are commonly used in embedded systems, game consoles, mobile phones and personal computers. For gamers, GPUs are stand alone cards plugged into a PCI Express bus. The GPU circuitry can also be on the CPU chip.

Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model.

Glove devised from Global Vectors, is an unsupervised learning algorithm for obtaining vector representation of words. This is achieved by mapping words into meaningful spaces so that the distance between them relates to semantic similarities. The algorithm is trained on the aggregated global co-occurrence statistics from a corpus and the results are substructures of the word vector space.

Gradient Clipping is a technique to prevent exploding gradients in very deep networks. A common way for gradient clipping is rescaling gradients so that their norm is at most a particular value. With gradient clipping, pre-determined gradient threshold is introduced, and then gradient norms that exceed this threshold are scaled down to match the norm.

Grid Search is a tuning technique that attempts to compute the optimum values of hyperparameters. It is an exhaustive search that is performed on a specific parameters values of a model.

GoogLeNet is a 27-layers CNN architecture constructed using the dimension-induced inception modules fitted linearly. It is 22-layers deep plus 5 pooling layers. A pretrained version of GoogLeNet can be loaded on either ImageNet or Places365 datasets.

Hidden layer is an ANN layer

located between the input and output layers of an algorithm, where it performs nonlinear transformations of the inputs in the network. The hidden layer applies weights to the input layers and produces an output through an activation function, therefore, is not visible as a network output.

Histogram is a data

visualization technique commonly used in ML that represents the frequency distribution of a continuous variable. Histograms plot data by splitting into intervals called bins whose areas are proportional to the corresponding frequencies.

Hierarchical clustering (also called

hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. It is best suited for hierarchical data, such as taxonomies. Strategies for hierarchical clustering generally fall into two types: Agglomerative, a "bottom-up" approach, and Divisive, a "top-down" approach.

Historical data is collected

data about past events and circumstances pertaining to a particular subject. It includes data generated either manually or automatically and it is commonly used as a basis for forecasting the future data or trends.

Hyperparameters are parameters

that are set before the learning process begins. These parameters affect how well a model trains. Unlike "model parameters", hyperparameters are "tunable". Some examples of hyperparameters in machine learning include number of epochs, learning rate, and number of clusters in a clustering algorithm.

Hyperplanes are decision

boundaries that help classify data points. Data points on either side of the hyperplane can be attributed to different classes. The dimension of the hyperplane depends upon the number of features. The most common example of hyperplanes in practice is with support vector machines.

Inclusiveness is one of the fundamental AI principles. It consists in creating AI systems that engage and empower people and in using inclusive design practices to eliminate unintentional barriers that could exclude people.

Inductive learning is a learning procedure of machine learning, during which the system tries to induce a "general rule" from a set of observed instances. In other words, we learn the model from raw data, the so-called training set.

Image recognition is a supervised learning task that identifies images and categorizes them in one of several predefined distinct classes. The field of study aimed at enabling machines with this ability is called computer vision. Some implementations of image recognition include security and surveillance, face recognition, object detection, and medical image analysis.

Instance is a single object from which an ML model is learnt or used for prediction. In ML, an instance is described by a number of attributes including feature vectors, class label etc. Some ML models use more complex instance representations like having a relation between instances or parts of instances.

Irreducible error is the error that can't be reduced by creating good models. It is a measure of the amount of noise in our data. It is called irreducible error because no matter how good we get at estimating the target function, we cannot reduce this error.

ImageNet is a large visual database designed for use in visual object recognition software research. More than 14 million images have been hand-annotated by the project to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided.

Jaccard index also known as the

Jaccard similarity coefficient, compares two finite sets to identify the shared and distinct elements. It is a measure that emphasizes the similarity, with a range from 0% to 100%, and is formally defined as the size of the intersection divided by the size of the union of the finite sets. The higher the percentage the more similar the two finite sets.

Jaccard distance is a measure

of dissimilarity between two finite sets. It is the complement of the Jaccard index and is obtained by subtracting the Jaccard index from 100%. The higher the percentage the less similar the two finite sets.

J-measure is an information

theoretic means of quantifying the information content of a rule and applies it to two rule induction methods: one where the rules are generated via the intermediate representation of a decision tree and one where rules are generated directly from examples.

Julia is an MIT certified free

open source, high level, high performance, dynamic programming language designed to excel at scientific and numerical computing. It is well known for its high speed and computational power. Julia can be used for specialized domains such as data science, machine learning, and visualization.

Junction tree algorithm is the

main algorithm in probabilistic graphical models. The name arises from the fact that before performing numerical computations, one has to transform the graph of the probabilistic graphical model into a tree with a set of properties that allow the efficient computation of posterior probabilities. The junction tree algorithm is also known as clique tree.

JupterLab is the next-

generation Interface of Jupyter Notebook. It is a web-based interactive development environment for Jupyter notebooks, code, and data. It is flexible and has an extensible modular architecture. It offers more of an IDE-like experience.

Jupyter notebook is an open-

source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. It supports several languages like Python, Julia, R etc. and is largely used for data analysis, data visualization, numerical simulation, statistical modeling, machine learning and much more.

K-fold cross-validation is an ML evaluation technique that involves splitting input data into K-subsets, training an ML model on all but one (K-1) subsets then evaluating the model on the subset not used for training. The process is repeated K-times, with a different subset excluded from training (and reserved for evaluation) each time. K-fold cross validation is useful for detecting overfitting.

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. It was developed with a focus on enabling fast experimentation and on being user-friendly, modular, and extensible. It is the most used deep learning framework among top-5 winning teams on Kaggle.

Kernel is a linear classifier for solving nonlinear problems. The process involves transforming linearly inseparable data instances to linearly separable ones using the kernel function. Through the function, the kernel maps the original nonlinear observations into higher-dimensional space in which they become separable.

K-means clustering is a method of

vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. It is popular for cluster analysis in data mining.

K-nearest neighbors algorithm

(KNN) is a supervised machine learning algorithm which can be used for both classification and regression problems. KNN uses 'feature similarity' to predict the value of new data points. It first identifies the k points in the training data that are closest to the test value and calculates the distance between all those categories. The test value will belong to the category whose distance is the least.

Layer is the highest-level building block in deep learning. A layer is a container that usually receives weighted input, transforms it with a set of mostly non-linear functions and then passes these values as output to the next layer. Some common types of deep learning layers include input layers, convolution and fully connected layers, activation layers, pooling layers, and output layers.

Labeled data is a designation for pieces of data that have been tagged with one or more labels identifying certain properties or characteristics, or classifications or contained objects. Labels make that data specifically useful in supervised machine learning.

L2-regularization is a technique that penalizes models to prevent overfitting, it adds an extra term to the cost function, a term called the regularization term.

Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models. it is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection.

Lemmatization is one of the most common text preprocessing techniques used in natural language processing and machine learning. It is a process of grouping words together with the same root word (lemma) but with different inflections or derivatives of meaning so they can be analyzed as one item. For example, to lemmatize "dogs", "dog's", and "dogs'" means taking away the suffixes "s", "'s", and "s'" to bring out the root word "dog".

LeNet is a convolutional neural network (CNN) structure proposed by Yann LeCun et al. in their 1998 paper. The authors' implementation of LeNet was used primarily for OCR and character recognition in documents. The LeNet architecture is straightforward and small in terms of memory footprint: it can even run on the CPU, which makes it perfect for teaching the basics of CNNs.

Libraries are collections of pre-written (and compiled) code that you can make use of in your own project. NumPy, Pandas & Jupyter are examples of libraries popularly used in data science, while Scikit-Learn, Apache Spark, TensorFlow, Keras & PyTorch are libraries specifically designed for deep learning and machine learning. Finally, libraries like Plotly, Matplotlib, Seaborn & Bokeh are popular for data visualization.

Linear Discriminant Analysis(LDA) is a linear transformation technique that is most commonly used for dimensionality reduction in the pre-processing step for pattern-classification and machine learning applications. It is a supervised learning algorithm and computes the directions ("linear discriminants") that will represent the axes that maximize the separation between multiple classes.

Linear regression is an ML supervised algorithm mostly used to find out the relationship between variables. Linear regression predicts values based on independent variables and these values are within a continuous range with a constant slope.

Logistic regression is an ML supervised classification algorithm for predicting the probability of a target feature. Logistic regression transforms its output using the logistic sigmoid function to return a probability value that can be mapped to two or more labels.

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). LSTMs are quite useful when our neural network needs to switch between remembering recent things, and things from a long time ago.

Loss function evaluates how well a specific algorithm models the given data, such that the larger the loss function metric, the higher the deviation of the predicted values from the actual value. With the help of optimization functions, the loss functions learn to reduce the error in prediction.

Learning rate is a hyper- parameter that defines the amount of minimization in the cost function in each iteration of a training process. In simple terms, the rate at which we descend towards the minima of the cost function is the learning rate. One should choose the learning rate very carefully since it should neither be very large that the optimal solution is missed and nor should be very low that it takes forever for the network to converge.

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+, and produces publication-quality figures in a variety of hardcopy formats and interactive environments across platforms.

Mean Squared Error (MSE) is a risk function computed as the average squared difference between the predicted and actual values. MSE tells you how close a regression line is to a set of points and it corresponds to the average of the squared error loss.

Machine learning is the science of getting computers to learn and act like humans do. As a data science technique, it is used to extract patterns from data; allowing computers to identify related data, and forecast future outcomes, behaviors, and trends. This learning is improved over time in an autonomous fashion by feeding data and information in the form of observations and real-world interactions.

MLOps is based on DevOps principles and practices that increase the efficiency of workflows. For example, continuous integration, delivery, and deployment. MLOps applies these principles to the machine learning process.

Managed services are services used to enhance the machine learning processes. Conventional ML setup takes time, and sometimes involves a fair amount of troubleshooting ensuring one has the right combination of software versions that are compatible with one another. With managed services for ML, a pre-optimized ready-made environment is provided for your machine learning development.

Mining is the process of extracting unknown patterns and hidden information from large quantities of data. Mining techniques assume that the unrevealed relationships exist within the data under examination and the actual mining task can be semi-automated or completely automated.

Modeling is the process of training a machine learning algorithm to predict the labels from the features, tuning it for the business need, and validating it on holdout data.

Machine Learning Model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data.

Markov Decision Process is a reinforcement learning model of predicting outcomes. It is a mathematical process to model decision-making in situations where outcomes are partly random and partly under the control of a decision-maker. It does not assume knowledge of an exact mathematical model. However, the Markov decision process incorporates the characteristics of actions and motivations.

Maximum Likelihood Estimation (MLE) is a probabilistic framework for solving the problem of density estimation. It involves maximizing a likelihood function in order to find the probability distribution and parameters that best explain the observed data and it provides a framework for predictive modeling in machine learning where finding model parameters can be framed as an optimization problem.

Multilayer Perceptron (MLP) is a

classical type of ANN that consists of an input layer, a hidden layer and an output layer. MLP utilizes a backpropagation technique, a supervised learning algorithm, for training and can distinguish data that is not linearly separable. It is suitable for both classification and regression problems.

Maximum Pooling or max pooling is a

sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. This is done to help overfitting and to reduce the computational cost by reducing the number of parameters to learn.

Model Registry lets you keep

track of all the models in your Azure Machine Learning workspace. Models are identified by name and version. Each time you register a model with the same name as an existing one, the registry assumes that it's a new version, you can also provide additional metadata tags and then use the tags when you search for models.

Multivariate Regression is a

technique that estimates a single regression model with more than one outcome variables. When there are more than one predictor variables in a multivariate regression model, the model is a multivariate multiple regression model. Multivariate regression analysis is not recommended for small samples and in order for it to make sense, the outcome variables should be at least moderately correlated.

Non-parametric Algorithms do not

make assumptions regarding the form of the mapping function between input data and output. Consequently, they are free to learn any functional form from the training data. A simple example is the K-nearest neighbors (KNN) algorithm. KNN does not make any assumptions about the functional form, but instead uses the pattern that points have similar output when they are close.

Normalization is the process of

transforming a piece of text into a canonical (official) form. For example, the verb "to be" may show up as "is", "am", "are", and so on. Or a document may contain alternative spellings of a word, such as "behavior" vs. "behaviour".

Normal distribution is the most

important distribution in statistics because it describes many natural random phenomena: heights, weights, ages, rainfall data, measurement errors etc. The normal distribution is a probability function that describes how the values of a variable are distributed. It is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. It is also known as bell curve or Gaussian distribution.

Natural Language Processing(NLP)

is the use of algorithms to determine properties of natural, human language so that computers can understand what humans have written or said. NLP includes teaching computer systems how to extract data from bodies of written text, translate from one language to another, and recognize printed or handwritten words. Virtual assistants — such as Siri— and chat bots are classic examples of NLP.

NaN Trap happens when one

number in the model becomes a NaN during training (e.g., when a value exceeds the floating-point precision limit), which causes —due to math operations— many or all other numbers in the model to eventually become a NaN.
NaN is an abbreviation for "Not a Number."

Neural Network is a model that is

inspired from the brain's architecture and is composed of layers (at least one of which is hidden) consisting of simple connected units or neurons followed by nonlinearities. Neural networks are designed to recognize patterns and relationships in data.

Neuron is a mathematical function that takes a group of weighted inputs, applies a nonlinear function known as an activation function and returns an output. Inputs to a neuron can either be features from a training set or outputs from a previous layer's neuron.

Non-max Suppression is a technique used in many computer vision algorithms. It is a class of algorithms to select one entity (e.g. bounding boxes) out of many overlapping entities. The selection criteria can be chosen to arrive at particular results. Most commonly, the criteria is some form of probability number along with some form of overlap measure.

Node is a computational unit that has either single or multiple weighted input connections, a transfer function combining the inputs and an output connection. Nodes are organized into layers to comprise a network.

NumPy is the fundamental package for scientific computing with Python. It is an open source general-purpose array-processing package which provides a high-performance multi-dimensional array object, and tools for working with these arrays. It can also be an efficient multi-dimensional container of generic data.

Notebooks are made up of one or more cells that allow for the execution of the code snippets or commands within those cells. The most common notebook in use today is the Jupyter Notebook. It supports several languages including Python, Julia, and R.

Neural Turing Machine is an RNN model that combines the fuzzy pattern matching capabilities of neural networks with the algorithmic strength of programmable computers. An NTM is designed to solve tasks that require writing to and retrieving information from external memory.

One-hot encoding is a process of converting categorical variables into binary vectors that are run in ML algorithms for accurate predictions. The process involves mapping the categorical values into integer values then each integer value is represented as a binary vector that is all zero values except the index of the integer which is marked with a 1.

Operationalization is the process of strictly defining variables into measurable factors. The process defines fuzzy concepts and allows them to be measured, empirically and quantitatively.

One-shot learning is a deep learning problem found mostly in computer vision. It applies object categorization algorithms on only one instance when training data then learns to re-identify that instance when testing data. A popular example of One-shot learning is the facial recognition system.

OpenCV is a library of programming functions mainly aimed at real-time computer vision. It was originally developed by Intel and later supported by Willow Garage before Itseez, which was later acquired by Intel. The library is cross-platform and free for use under the open-source BSD license.

Out-of-core refers to processing data that is too large to fit into a computer's main memory. The algorithms that are designed to process this kind of data are called external memory algorithms or out-of-core algorithms and must be optimized to efficiently fetch and access data stored in slow bulk memory or when memory is on a computer network.

Overfitting occurs when the machine learning algorithm learns a model that matches the training data so closely that the model fails to make correct predictions on new data. The problem of overfitting is usually solved by regularization, reducing the model complexity, or early stopping.

Outlier is any data point which deviates significantly from other observations in a dataset. It can be caused by measurements or execution error. In other words, an outlier is an observation that diverges from an overall pattern on a sample. The analysis of outlier data is referred to as outlier analysis or outlier mining.

Optimization is the selection of a best element from some set of available alternatives. It is an essential ingredient of ML algorithms. It starts with defining some loss/cost function and ends with minimizing it using an optimization technique. The choice of optimization algorithm can make a difference between getting good accuracy in hours or days.

Padding is a term relevant to convolutional neural networks as it refers to the amount of pixels added to an image when it is being processed by the kernel of a CNN. For example, if the padding in a CNN is set to zero, then every pixel value that is added will be of value zero. If, however, the zero padding is set to one, there will be a one pixel border added to the image with a pixel value of zero.

Principal Component Analysis(PCA) is a dimensionality-reduction technique that is based mostly on exact mathematical calculations and is used in exploratory data analysis and for making predictive models. It consists in reducing the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Privacy and Security are two of the main principles that should characterize AI systems. This basically means that every deployed AI system should be secure and respect existing privacy laws, be transparent about data collection, use good controls and de-identification techniques, and have policies that facilitate access to the data that the AI system needs to operate effectively.

Pooling is commonly introduced between convolution layers. This is basically done to reduce a number of parameters and prevent overfitting. The pooling operation involves sliding a two-dimensional filter over each channel of feature map and summarizing the features lying within the region covered by the filter. The two common functions used in the pooling operation are average pooling and maximum (max) pooling.

Precision is a metric for binary classification (Yes/No) models. It measures the model's performance at classifying positive observations (i.e. "Yes"). Precision (P) can be computed as the number of True Positives (TP) divided by the sum of True Positives and False Positives (FP) i.e. $P=TP/(TP+FP)$.

Part-of is a feature engineering technique which aims at extracting new data from already existing features.

Perceptron is an algorithm used for supervised learning of binary classifiers. Binary classifiers decide whether an input, usually represented by a series of vectors, belongs to a specific class. In short, a perceptron is a single-layer neural network. It consists of four main parts including input values, weights and bias, net sum, and an activation function.

Pandas is an open-source Python library designed for analyzing and manipulating data. It is particularly good for working with tabular data and time-series data.

Plotly is a data plotting library with a clean interface designed to allow you to build your own APIs. It allows users to import, copy and paste, or stream data to be analyzed and visualized. For analysis and styling graphs, Plotly offers a Python sandbox, datagrid, and GUI. Python scripts can be saved, shared, and collaboratively edited in Plotly. It is also supported on other programming languages including R and Julia.

Prophet is an open-source library developed by Facebook and designed for automatic additive time series forecasting which supports non-linear trends with seasonality and holiday effects. It works best for time series data with strong seasonal effects and multiple seasons of historical data.

PyTorch is a free open-source ML library based on the Torch library released under the modified BSD license. Developed by Facebook's AI research lab, PyTorch is a strong player in artificial intelligence and deep learning and is primarily used for computer vision and NLP.

Python is a high-level open source coding language known for object-oriented, interpreted and interactive programming purposes. It combines remarkable power with clear readable syntax thus convenient for data science and writing system scripts. Python is named after Monty Python and is not some complicated tech metaphor or a complicated acronym.

Presentation of data refers to the organization of data into tables, graphs or charts in an attractive and useful manner such that it can be easily interpreted, and logical and statistical conclusions can be derived from the collected measurements. The three main forms of presentation of data are: textual, tabular & graphical.

Parametric ML Algorithms make assumptions about the mapping function and have a fixed number of parameters. No matter how much data is used to learn the model, this will not change how many parameters the algorithm has. With a parametric algorithm, the form of the function is selected and then its coefficients learnt using the training data.

Pipeline is an independently executable workflow of a complete machine learning task. A pipeline should include machine learning tasks such as data preparation, training, validation, and deployment. There are two basic types of pipeline stages: Transformer and Estimator. A Transformer takes a dataset as input and produces an augmented dataset as output. An Estimator receives the output from a transformer to produce a model.

Quantile Transforms are a technique for transforming numerical input or output variables to have a Gaussian or uniform probability distribution. Many machine learning algorithms perform better when numerical variables have a Gaussian or standard probability distribution.

Q-learning is a basic form of Reinforcement Learning which uses Q-values (also called action values) to iteratively improve the behavior of a learning agent. The 'q' in q-learning stands for quality. Quality in this case represents how useful a given action is in gaining some future reward.

Query is a request for data or information from a database table or combination of tables. This data may be generated as results returned by Structured Query Language (SQL) or as pictorials, graphs or complex results, e.g., trend analyses from data-mining tools. For a machine to understand the query, it must be written according to a code known as query language.

Q-table is a simple lookup table for calculating the maximum expected future rewards for action at each state. In a Q-table, the columns are the actions and the rows are the states. There are four actions (up, down, left & right) in each state and the Q-table guides to the best action at each state.

Quartile is a type of quantile which divides the number of data points into four more or less equal parts, or quarters. The first quartile (Q_1) is defined as the middle number between the smallest number and the median of the data set. The second quartile (Q_2) is the median of a data set and 50% of the data lies below this point. The third quartile (Q_3) is the middle value between the median and the highest value of the data set.

Quantiles are cut points dividing the range of a probability distribution into continuous intervals with equal probabilities, or dividing the observations in a sample in the same way. There is one fewer quantile than the number of groups created.

Random Forest is a supervised learning algorithm. The random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. In general, random forests outperform decision trees.

R is a free programming language and environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS, and provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering and more) and graphical techniques. It is highly extensible via packages.

Real-time inferencing is the process of generating machine learning predictions in real time upon request. Typically, these predictions are generated on a single observation of data at runtime. Some examples of applications that can benefit from real-time inferencing are augmented reality, virtual reality, human-computer interfaces, self-driving cars, and any consumer facing apps that allow users to query models in real time.

Recommender system is the most successful and widespread commercial application of machine learning technologies in business. It is a subclass of information filtering system, either collaborative filtering or content-based system, that predicts the rating or preference of a user. Playlist generators, content recommenders, and product recommenders are common application areas of recommender systems.

Regression is an ML algorithm that allows prediction from data by learning the relationship between features of a given data and some observed, continuous-valued response. It is used in numerous applications ranging from predicting stock prices to understanding gene-regulatory networks. Different regression models differ based on the relationship between the dependent and independent variables and the number of independent variables being used.

Recall also known as sensitivity, is a measure of how a model correctly identifies True Positives (TP). It is the fraction of the number of positive predictions that were correctly identified from the total positives i.e.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.
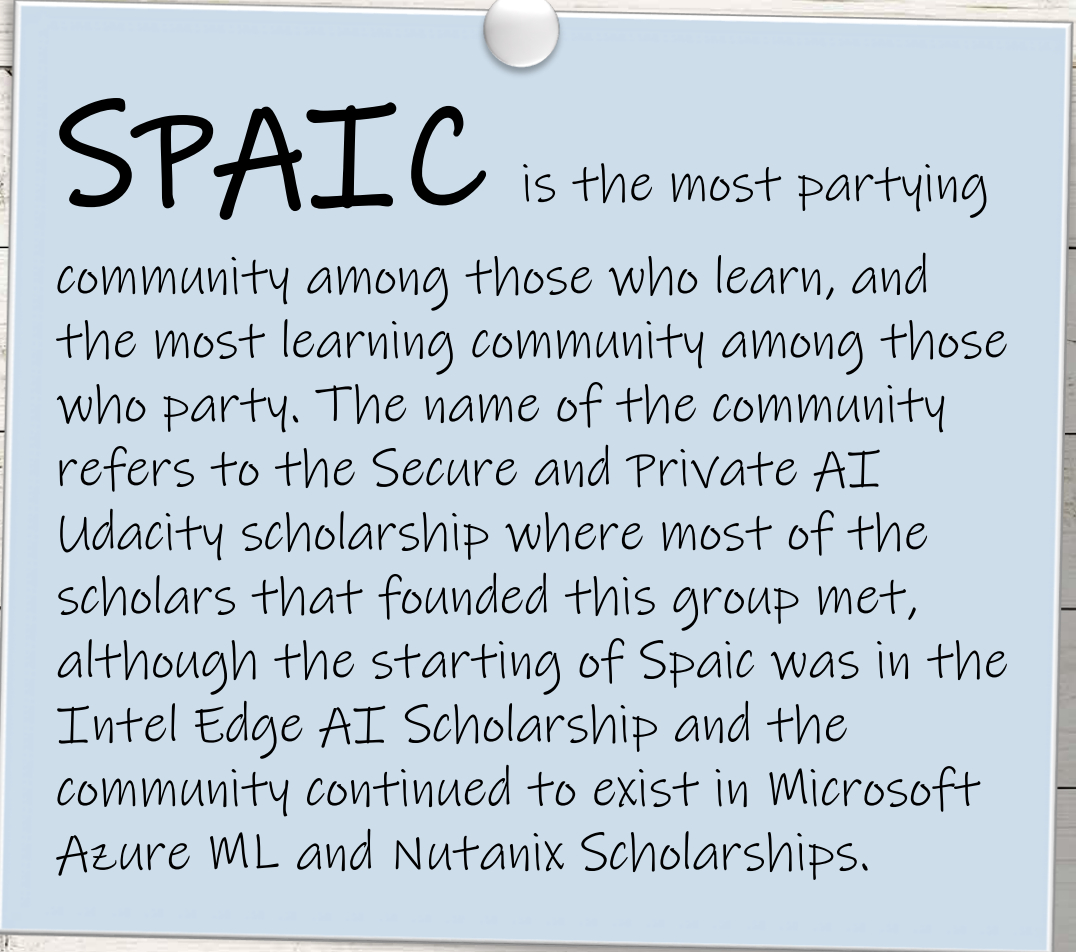
Reliability and Safety is an AI principle that aims to ensure that AI systems reliably operate in accordance with their intended purpose throughout their lifecycle. This includes ensuring AI systems are reliable, accurate and reproducible as appropriate. AI systems should not pose unreasonable safety risks, and should adopt safety measures that are proportionate to the magnitude of potential risks.

Rectified Linear Unit (ReLU) is a commonly used activation function in deep neural networks. The function returns 0 if it receives any negative input, but for any positive value k, it returns that value back. The major benefit of using ReLU is that it has a constant derivative value for all inputs greater than 0. The constant derivative value helps the network to train faster.

Receiver Operating Characteristic(ROC) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: true positive and false positive rates. A straight line represents the ROC curve of a purely random classifier; a good classifier has its curved line as far from the straight line as possible. The Area Under the Curve (AUC) represents the region within these two lines; in real life, it falls between 0.5 and 1. A perfect prediction would have an AUC of 1.

Run in Microsoft Azure is a single execution of training script. It contains all artifacts associated with a training process: output files, metrics, logs, and a snapshot of the directory that contains your scripts. You produce a run when you submit a script to train a model. Before you create a new run, you must first create an experiment.

SPAIC is the most partying community among those who learn, and the most learning community among those who party. The name of the community refers to the Secure and Private AI Udacity scholarship where most of the scholars that founded this group met, although the starting of Spaic was in the Intel Edge AI Scholarship and the community continued to exist in Microsoft Azure ML and Nutanix Scholarships.

Scatter Plot is a graph of two sets of data along the two axes. It is used to visualize the relationship between two variables. Each row in the data table is represented by a marker in the scatter plot whose position depends on its values in the columns set on the X and Y axes.

Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

Semi-supervised Learning is intermediate between supervised learning and unsupervised learning. Supervised learning starts with training data that are tagged with the correct answers (target values). After the learning process, you wind up with a model with a tuned set of weights, which can predict answers for similar data that haven't already been tagged.

Seaborn is a data visualization library for making statistical graphics in Python. It uses Matplotlib underneath and integrates closely with pandas data structure to generate simple interface and aesthetically better-looking plots of random distributions.

Scoring also known as prediction, is the process of generating values based on a trained machine learning model, given some new input data. The values or scores that are created can represent predictions of future values, but they might also represent a likely category or continuous outcome. The meaning of the score depends on the type of data you provide, and the type of model that you created.

Specificity is defined as the proportion of actual negative cases that was predicted as negative by our model. In simple terms, it is the ratio of true negatives to all negatives. Specificity is also known as true negative rate. It has the formula:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives + False Positives}}$$

Sentiment Analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Similarity Learning is an area of supervised machine learning in artificial intelligence. It is closely related to regression and classification, but the goal is to learn a similarity function that measures how similar or related two objects are. It has applications in ranking, recommendation systems, visual identity tracking, face verification, and speaker verification.

Softmax is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1.

Spark (Apache Spark) originally developed at UC Berkeley in 2009, is an open source data processing engine for large datasets. It is designed to deliver the computational speed, scalability, and programmability required for Big Data - specifically for streaming data, graph data, machine learning, and AI applications. It includes APIs from programming languages that are popular among data analysts and data scientists including Scala, Java, Python, and R.

Standardization (or Z-score normalization) is the process where the features are rescaled so that they will have the properties of a standard normal distribution with $\mu=0$ and $\sigma=1$, where μ is the mean (average) and σ is the standard deviation from the mean.

Stemming is the process of producing morphological variants of a root/base word. Stemming programs are commonly referred to as stemming algorithms or stemmers. A stemming algorithm reduces the words "chocolates", "chocolatey", "choco" to the root word, "chocolate" and "retrieval", "retrieved", "retrieves" are reduced to the stem "retrieve". Stemming is an important part of the pipelining process in Natural Language Processing.

Splitting Data is a machine learning practice done to avoid overfitting and model selection bias. The most common practice is splitting the dataset into three: training, validation, and test. This allows you to tune various parameters of the algorithm without making judgements that specifically conform to training data. Data split may be in the ratio of 60:20:20, 98:1:1, or 99:0.5:0.5 respectively depending on the size of the data.

Stacking is an ensemble learning algorithm. It learns heterogeneous weak learners in parallel and combines them by training a meta-model to output a prediction based on the different weak model predictions. The benefit of stacking is that it can harness the capabilities of a range of well-performing models on a classification or regression task and make predictions that have better performance than any single model in the ensemble. Stacking reduces the model bias.

Sensitivity is defined as the proportion of actual positive cases that was predicted as positive by our model. In simple terms, it is the ratio of true positives to all positives. Sensitivity is also known as recall or true positive rate. It has the formula:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Support Vector Machine (SVM)

is a classification method in supervised learning that seeks to use support vectors (cases close to the boundary) to find an optimal hyperplane separating items from different classes.

Storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion. It is considered the most effective way to share business information and drive outcomes.

Supervised Learning is a subclass of machine learning in which a desired model predicts the label for yet-unseen data. Supervised machine learning is analogous to a student learning a subject by studying a set of questions and their corresponding answers. After mastering the mapping between questions and answers, the student can then provide answers to new (never-before-seen) questions on the same topic.

Temporal Convolutional Network (TCN) is a framework which employs casual convolutions and dilations so that it is adaptive for sequential data with its temporality and large receptive fields.

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

Test Set is a dataset used to provide an unbiased evaluation of a final model fit on the training dataset. If the data in the test dataset has never been used in training (for example in cross-validation), the test dataset is referred to as a holdout dataset.

Training is the process of determining the ideal parameters comprising a model for a given training data set. For a linear model, this means finding the weights. For a tree, it involves identifying the split points.

Tuning also known as hyperparameter optimization, is the process of choosing an optimal hyperparameter for a learning algorithm to maximize a model's performance without overfitting. The hyperparameter value is used as a dial or knob to control the learning process.

T-test compares the averages of two populations to determine how different they are from each other. The test generates a T-score and P-value, which quantify exactly how different each population is and the likelihood that this difference can be explained by chance or sampling error.

Training Set or learning set is the subset of the dataset used to train a model. A training set can be structured in different ways. For sequential decision trees and those types of algorithms, it would be a set of raw text or alphanumeric data that gets classified or otherwise manipulated. For convolutional neural networks that have to do with image processing and computer vision, the training set is often composed of large numbers of images.

T-SNE (t-Distributed Stochastic Neighboring Entities) is a nonlinear dimensionality reduction machine learning algorithm for visualization based on Stochastic Neighbor Embedding. It is well-suited for embedding high-dimensional data for visualization in a low-dimensional space in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability.

Temporal Convolutional Network (TCN) is a framework which employs casual convolutions and dilations so that it is adaptive for sequential data with its temporality and large receptive fields.

Transfer Learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. In transfer learning, we take the pre-trained weights of an already trained model (one that has been trained on millions of images belonging to 1000's of classes, on several high power GPU's for several days) and use these already learned features to predict new classes.

Support Vector Machine (SVM)

is a classification method in supervised learning that seeks to use support vectors (cases close to the boundary) to find an optimal hyperplane separating items from different classes.

Storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion. It is considered the most effective way to share business information and drive outcomes.

Supervised Learning is a subclass of machine learning in which a desired model predicts the label for yet-unseen data. Supervised machine learning is analogous to a student learning a subject by studying a set of questions and their corresponding answers. After mastering the mapping between questions and answers, the student can then provide answers to new (never-before-seen) questions on the same topic.

Underfitting occurs when a

statistical model or machine learning algorithm cannot capture the underlying trend of the data. In this case, the resulting model has "not learned enough" from the training data, resulting in low generalization and unreliable predictions. Hence, the model can neither model the training data nor generalize to new data. An underfit machine learning model is not a suitable model.

Unintentional Bias is often defined as

prejudice or unsupported judgments in favor of or against one thing, person, or group as compared to another, in a way that is usually considered unfair. It is one of the several challenges that modern AI brings with it. One relevant example was with a major online advertising system that showed an ad for high-income jobs to men more often than it showed the ad to women.

Unit (AKA activation function) is a feature of an ANN that decides if a neuron should be activated or not. An example of a unit is the standard integrated circuit that can be "ON" (1) or "OFF" (0) depending on the inputs and which is similar to the behavior of linear perceptron in neural networks.

Unlabeled Data is a designation for pieces of data that have not been tagged with labels identifying characteristics, properties or classifications. Unlabeled data is typically used in various forms of machine learning.

Unsupervised Learning is an ML technique that mainly deals with unlabeled data. It allows the model to work on its own to discover patterns and information that was previously undetected. Also known as self-organization, unsupervised learning allows for modeling of probability densities over inputs.

Vectorization is a technique by which you can make your code execute fast. It is a very interesting and important way to optimize algorithms when you are implementing it from scratch. vectorization turns a piece of text into a vector. A vector is simply an array of numbers.

Validation is a process used, as part of training, to evaluate the quality of a machine learning model using the validation set. Because the validation set is disjoint from the training set, validation helps ensure that the model's performance generalizes beyond the training set.

Visualization of Data 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.

Voilà (Jupyter) is a new Python package and a recent addition to the Jupyter Notebook ecosystem, which can turn any Jupyter notebook into standalone web applications. It allows conversion of Jupyter Notebooks into interactive dashboards that can be shared with others. It is secure and customizable thus, allowing control over what readers experience.

Variance Error is an error from sensitivity to small fluctuations in the training dataset estimated by a machine learning algorithm. High variance can cause an algorithm to model the random noise in the training data rather than the intended outputs thus overfitting.

Vanishing Gradient is a problem that

happens in very deep neural networks, typically recurrent neural networks, that use activation functions whose gradients tend to be small. Because these small gradients are multiplied during backpropagation, they tend to "vanish" throughout the layers, preventing the network from learning long-term dependencies. Common ways to counter this problem is to use activation functions like ReLU or LSTM that do not suffer from small gradients.

Voronoi Diagram is a diagram created

by scattering points at random on a Euclidean plane. It is the computational geometry concept that represents partition of the given space into regions, with bounds determined by distances to a specified family of objects. In machine learning, Voronoi diagrams are used to do 1-nearest neighbor classifications.

WaveNet is a deep autoregressive model for generating human-like voices, where raw audio is fed as input to the model then it takes the speech synthesis to another level. The technique generates relatively realistic humanlike voices by directly modelling waveforms using a neural network method.

Weaponization refers to the fact that machine learning is making the creation of malware easier, faster and cheaper, meaning that organizations should prepare for an upsurge in the volume and sophistication of cyberattacks.

Weights are learnable parameters of a machine learning model. Weights control the signal (or the strength of the connection) between two neurons. In other words, a weight decides how much influence the input will have on the output.

Workspace - **Azure** is the top-level resource for Azure Machine Learning, providing a centralized place to work with all the artifacts you create when you use Azure Machine Learning. The workspace keeps a history of all training runs, including logs, metrics, output, and a snapshot of your scripts.

Wrangling of Data is the process of cleaning and transforming data to make it more appropriate for data analysis. The process generally follows these main steps:
a) exploring the raw data and checking the general quality of the dataset;
b) transforming the raw data; and
c) validating and publishing the data.

Word2Vec is one of the most popular techniques to learn word embeddings using shallow neural networks. It is a statistical method for efficiently learning a standalone word embedding from a text corpus. Word2Vec was developed by Tomas Mikolov, et al. at Google in 2013 as a response to make the neural-network-based training of the embedding more efficient and since then has become the de facto standard for developing pre-trained word embedding.

Xplainability is the degree to which the inner workings of the model can be examined, understood, and explained. One of the challenges with Machine Learning models is their opacity. This opacity depends a lot on the class of algorithms used to train the model. Decision Trees produce the lowest opacity trained models which are essentially self-explanatory, while Deep Neural Networks tend to produce the most opaque models that are quite difficult to understand and explain even for the experts.

Xperiment is a logical container in an Azure ML Workspace that helps you organize the model training process. It hosts run records which can include run metrics and output artifacts from your experiments.

XGBoost is an open-source software library which provides a gradient boosting framework for C++, Java, Python, R, Julia, Perl, and Scala. It aims to provide a "Scalable, Portable and Distributed Gradient Boosting (GBM, GBRT, GBDT) Library". It runs on a single machine, as well as the distributed processing frameworks Apache Hadoop, Apache Spark, and Apache Flink.

Xploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. This can be some kind of readable format like an Excel spreadsheet or a complex visual model that visualizes data points. An in-depth exploration means structuring and visualizing data so that patterns, outliers, anomalies, and other factors can be identified.

Xtractive Sentence Summarization picks a sentence directly from a document based on a scoring function to create a coherent summary. The flow of extractive sentence summarization involves finding salient content by computing TF metrics, scoring the sentences based on the representation and producing a summary based on K-most important sentences.

Yarowsky Algorithm is an unsupervised learning algorithm that uses the collocation and discourse properties of human languages for word sense disambiguation. The algorithm consists of two loops; the inner loop which is a supervised learning algorithm and the outer loop which is guided by a seed set of rules to start with then assign labels to unlabeled data in each iteration.

YOLO is an acronym standing for the phrase "You Only Look Once" and refers to a fast object detection algorithm. Previous attempts at building object detection algorithms involved running object detectors or object localizers multiple times over a single image. Instead of needing multiple executions over a single image, YOLO detects objects by sending an image through a single forward pass through a convolutional neural network.

Zero-Shot Learning is an ML

setup problem where during testing, a classifier assesses samples from classes that were not trained and predicts the category where they belong to. ZSL is widely studied in computer vision, NLP and machine perception and can be viewed as an extreme case of domain adaptation.

Z-test is a family of statistical

tests that assumes a normal distribution model to compute test statistics. It is used to determine whether two population means are different when the variances are known and the sample size is large. The standard deviation should be known for an accurate z-test to be performed.

Z-score also called a standard

score gives you an idea of how far from the mean a data point is. But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

Zero-padding, in convolutional

neural networks, refers to surrounding a matrix with zeros. This helps preserve features that exist at the edges of the original matrix and control the size of the output feature map.