Machine Learning

Tushar B. Kute,
http://tusharkute.com

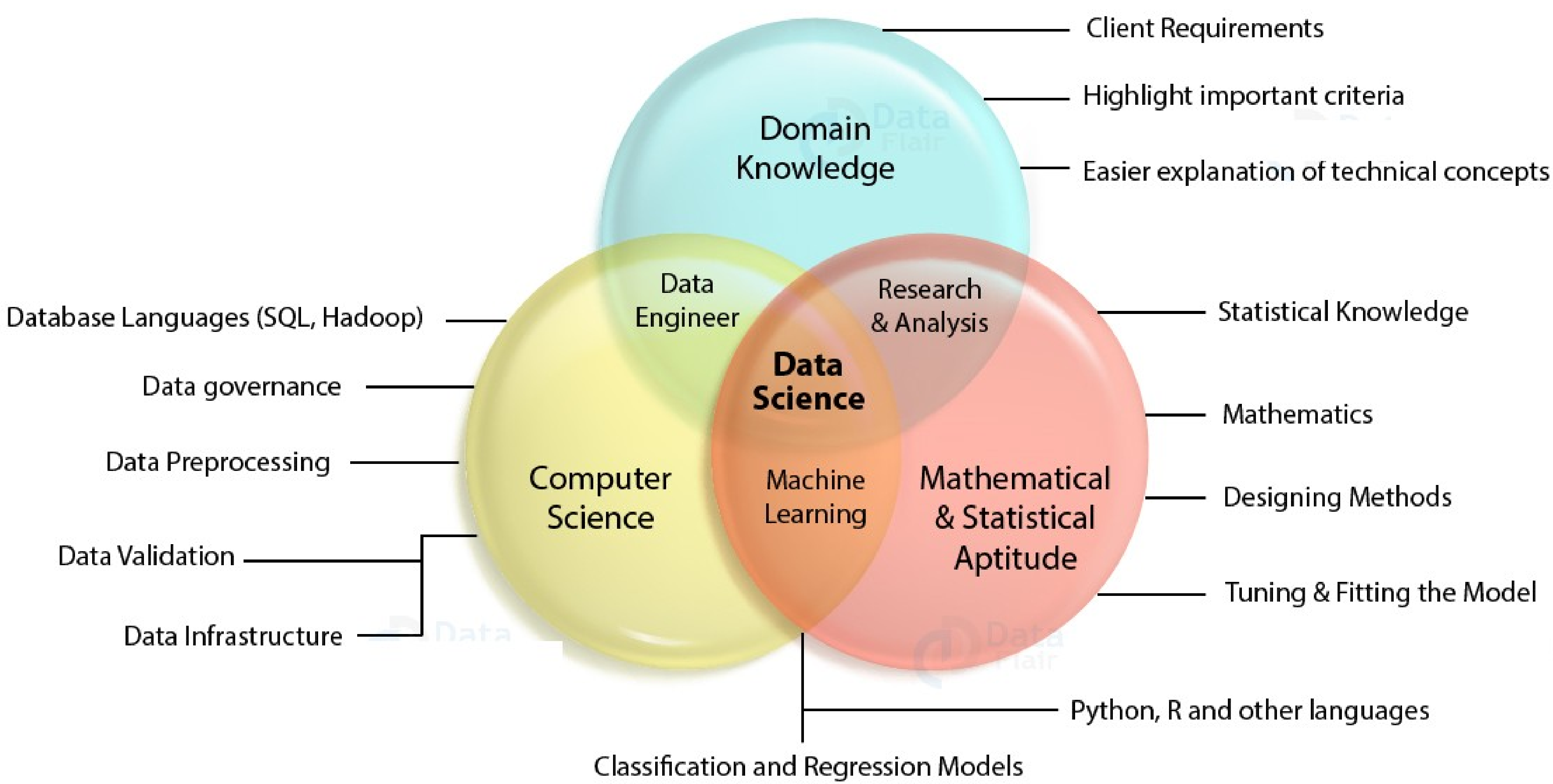
What is Data Science ?

- Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.
- Data science is related to data mining, machine learning and big data.
- Data science (DS) is a multidisciplinary field of study with goal to address the challenges in big data.
- Data science principles apply to all data – big and small.

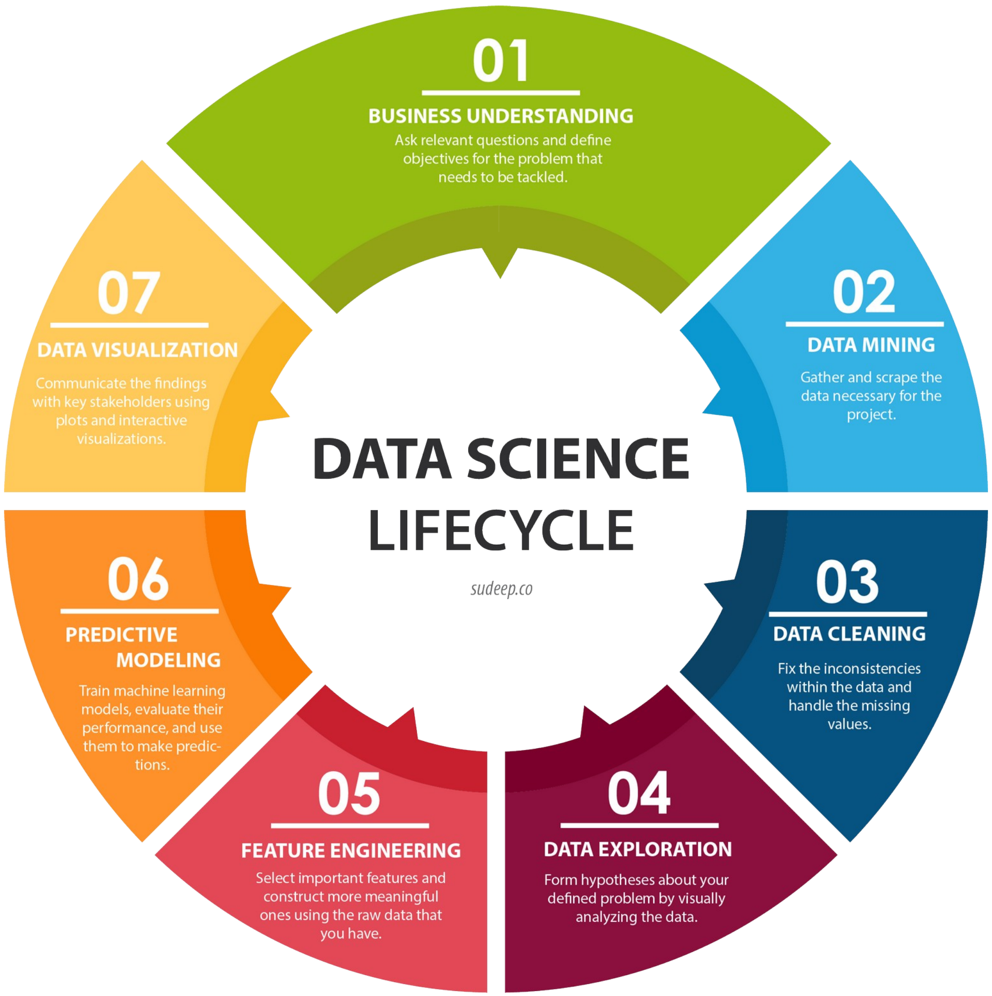
What is Data Science ?

- Theories and techniques from many fields and disciplines are used to investigate and analyze a large amount of data to help decision makers in many industries such as science, engineering, economics, politics, finance, and education
  - Computer Science
    - Pattern recognition, visualization, data warehousing, High performance computing, Databases, AI
  - Mathematics
    - Mathematical Modeling
  - Statistics
    - Statistical and Stochastic modeling, Probability.

Data Science Requirements



Data Science Life-Cycle



Machine Learning

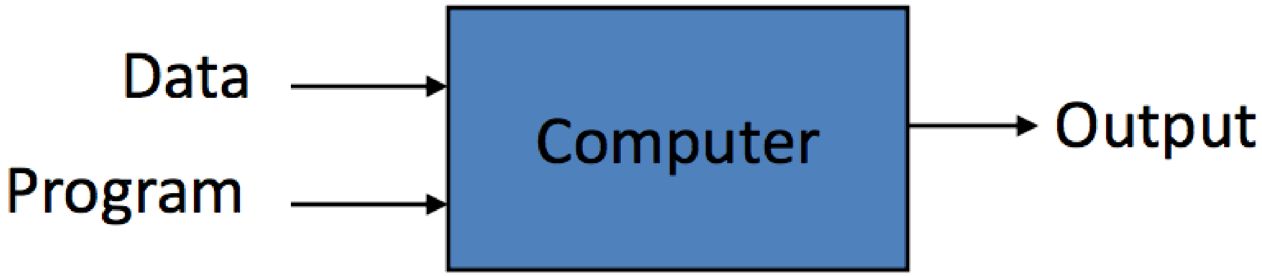
- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.
- The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.
- The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Origins of Machine Learning

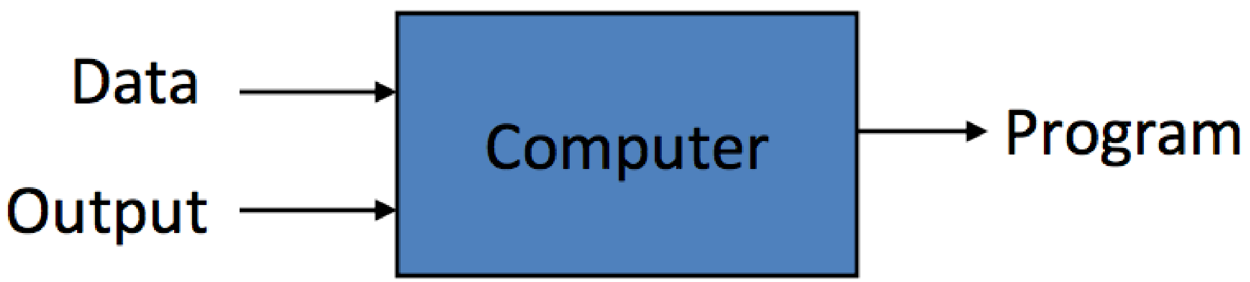
- The earliest databases recorded information from the observable environment.
- Astronomers recorded patterns of planets and stars; biologists noted results from experiments crossbreeding plants and animals; and cities recorded tax payments, disease outbreaks, and populations. Each of these required a human being to first observe and second, record the observation.
- Today, such observations are increasingly automated and recorded systematically in ever-growing computerized databases.

Machine Learning

Traditional Programming

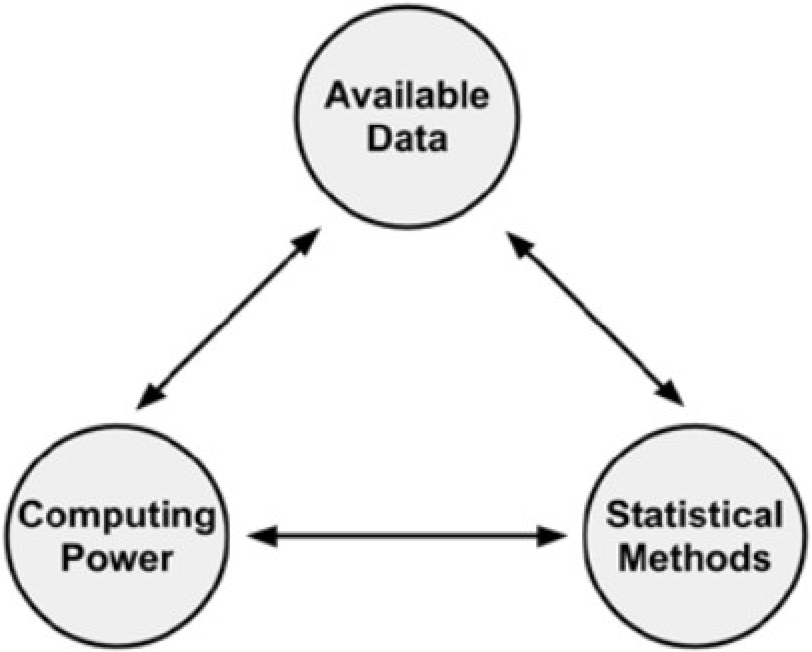


Machine Learning



Machine Learning

- The field of study interested in the development of computer algorithms for transforming data into intelligent action is known as machine learning.



Data Mining

- A closely related sibling of machine learning, data mining, is concerned with the generation of novel insight from large databases (not to be confused with the pejorative term "data mining," describing the practice of cherry-picking data to support a theory).
- Although there is some disagreement over how widely the two fields overlap, a potential point of distinction is that machine learning tends to be focused on performing a known task, whereas data mining is about the search for hidden nuggets of information.

Timeline

1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.

1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.

1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.

1967
The Nearest Neighbor Algorithm was written.

1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.

1997
IBM's Deep Blue beats the world champion at Chess.

2002
A software library for Machine Learning, named Torch is first released.

2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go competition.

Well-Posed Learning Problems

- A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Well-Posed Learning Problems - Examples

- A checkers learning problem
  - Task T : playing checkers
  - Performance measure P : percent of games won against opponents
  - Training experience E : playing practice games against itself
- A handwriting recognition learning problem
  - Task T : recognizing and classifying handwritten words within images
  - Performance measure P : percent of words correctly classified
  - Training experience E : a database of handwritten words with given classifications

Well-Posed Learning Problems - Examples

- A robot driving learning problem
  - Task T : driving on public four-lane highways using vision sensors
  - Performance measure P : average distance traveled before an error (as judged by human overseer)
  - Training experience E : a sequence of images and steering commands recorded while observing a human driver

Well-Posed Learning Problems - Examples

- A robot driving learning problem
  - Task T : driving on public four-lane highways using vision sensors
  - Performance measure P : average distance traveled before an error (as judged by human overseer)
  - Training experience E : a sequence of images and steering commands recorded while observing a human driver

Uses and Abuses

- Predict the outcomes of elections
- Identify and filter spam messages from e-mail
- Foresee criminal activity
- Automate traffic signals according to road conditions
- Produce financial estimates of storms and natural disasters
- Examine customer churn
- Create auto-piloting planes and auto-driving cars
- Identify individuals with the capacity to donate
- Target advertising to specific types of consumers

Case Study

mitu skillologies

सकाळ

विद्यापीठात विद्यार्थ्यांचा 'एक्झिट पोल'

'रँडम फॉरेस्ट मॉडेल'नुसार युतीच राज्यात आघाडीवर

पुणे, ता. २१ : राज्यात भाजप आणि शिवसेना युती आघाडीवर असेल, असा अंदाज वर्तविणाऱ्या चाचण्यांचे कल (एक्झिट पोल) नुकतेच प्रसिद्ध झाले आहेत. सावित्रीबाई फुले पुणे विद्यापीठातील विद्यार्थ्यांनीही त्याला दुजोरा दिला आहे. भारतीय जनता पक्षाला १७ ते २३ आणि शिवसेनेला १६ ते २१ जागा मिळतील, असा अंदाज विद्यार्थ्यांनी 'रँडम फॉरेस्ट मॉडेल' पद्धत वापरून वर्तविला आहे. राष्ट्रवादी काँग्रेसला ३ ते ९ व काँग्रेसला १ ते ६ जागा मिळतील, असा अंदाज त्यांनी वर्तवला आहे.

विद्यापीठाच्या संख्याशास्त्र विभागातील एमएस्सी (द्वितीय वर्ष)

करणारे विनय तिवारी, आर. विश्वनाथ, शरद कोळसे या विद्यार्थ्यांनी सहायक प्राध्यापक डॉ. आकांक्षा काशीकर यांच्या मार्गदर्शनाखाली हा अंदाज दिला आहे.

निवडणूक आयोगाच्या संकेतस्थळावरून सर्वेक्षणासाठी लागणारी माहिती त्यांनी मिळविली. जनमानसाचा कल ओळखण्यासाठी 'सीएसडीएस-लोकनीती' सर्वेक्षण अहवालातून नोंदी घेतल्या. त्याचबरोबर सध्याच्या सरकारच्या कामगिरीबद्दल लोकांच्या प्रतिक्रिया, पंतप्रधानपदाच्या संभाव्य उमेदवारांची लोकप्रियता, मागील निवडणुकीतील आपले मत यंदा बदलू इच्छिणारे मतदार यांचा अभ्यास करण्यात आला. या अंदाजासाठी रँडम फॉरेस्ट मॉडेल वापरण्यापूर्वी २००९ आणि २०१४च्या निवडणुकांचे अंदाज पडताळून पाहण्यात आले. हे अंदाज प्रत्यक्ष निकालांशी पडताळून पाहिले असता, ते जवळपास ९६ टक्के जुळत असल्याचे निदर्शनास आले. म्हणूनच अभ्यासात माहितीच्या विश्लेषणासाठी या पद्धतीचा वापर करण्यात आला, असे डॉ. काशीकर यांनी सांगितले.

> संख्याशास्त्र आणि संगणकशास्त्र याची सांगड घालून आणि मशिन लर्निंगच्या साह्याने उपलब्ध माहितीचे विश्लेषण केले. संख्याशास्त्रातील अभ्यासाची वेगवेगळी मॉडेल वापरून १९७७ पासून ते आतापर्यंतच्या लोकसभा आणि विधानसभा निवडणुकीतील माहितीचा अभ्यास केला. त्यामुळे संख्याशास्त्राचा वापर करून वर्तविलेला अंदाज हा निवडणुकीच्या निकालांच्या जवळ जाणारा असेल.
>
> – शरद कोळसे, विद्यार्थी

Pune, Main
22/05/2019 Page No. 4

Recognizing patterns

- A machine learning algorithm takes data and identifies patterns that can be used for action.
- In some cases, the results are so successful that they seem to reach near-legendary status.

How do machine learn ?

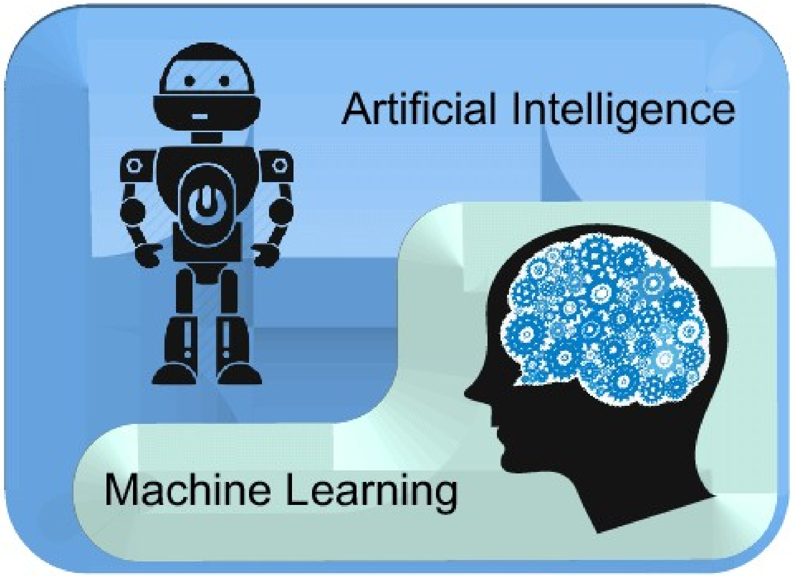
- A commonly cited formal definition of machine learning, proposed by computer scientist Tom M. Mitchell, says that a machine is said to learn if it is able to take experience and utilize it such that its performance improves up on similar experiences in the future.
- This definition is fairly exact, yet says little about how machine learning techniques actually learn to transform data into actionable knowledge.

Learning vs. Designing

- Artificial intelligence and machine learning are the part of computer science that are correlated with each other.
- These two technologies are the most trending technologies which are used for creating intelligent systems.
- Although these are two related technologies and sometimes people use them as a synonym for each other, but still both are the two different terms in various cases.

Learning vs. Designing

- AI is a bigger concept to design intelligent machines that can simulate human thinking capability and behavior, whereas, machine learning is an application or subset of AI that allows machines to learn from data without being programmed explicitly.



Artificial Intelligence

- Artificial intelligence is a field of computer science which makes a computer system that can mimic human intelligence.
- It is comprised of two words "Artificial" and "intelligence", which means "a human-made thinking power."
- Hence we can define it as,
  - Artificial intelligence is a technology using which we can create intelligent systems that can simulate human intelligence.

In short...

- AI leads to intelligence or wisdom.
- ML leads to knowledge.

Training a dataset

- The process of fitting a particular model to a dataset is known as training.
- Why is this not called learning? First, note that the learning process does not end with the step of data abstraction.
- Learning requires an additional step to generalize the knowledge to future data.
- Second, the term training more accurately describes the actual process undertaken when the model is fitted to the data.

Practical Machine Learning

Inputs

X	Y	Z
5	2	14
8	5	22
4	8	14
9	2	20
7	1	15
7	8	23

Output

Z = ?   ---> ML Model

Practical Machine Learning

X	Y	Z	Pre	Error
5	2	14	12	-2
8	5	22	21	-1
4	8	14	16	+2
9	2	20	20	0
7	1	15	15	0
7	8	23	22	-1

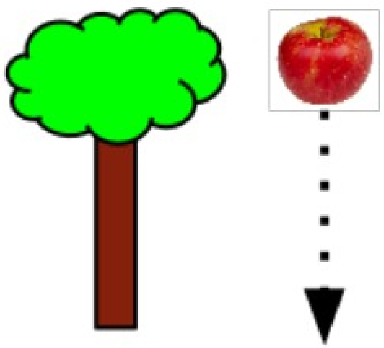
Z = 2X + Y ---> ML Model

Prediction ---> X = 6 Y = 8 Z = ?

if 20 == 19: 95%

Training a dataset

Observations → **Data** → **Model**



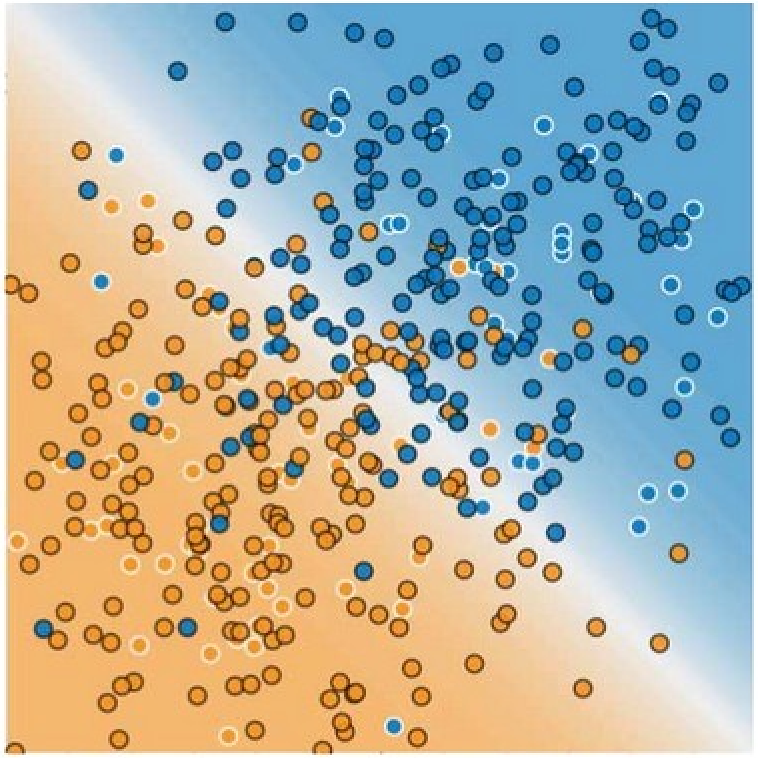
velocity	time
9.8	1
39.2	2
88.2	3
156.8	4
245	5

$$g = 9.8\ m/s^2$$

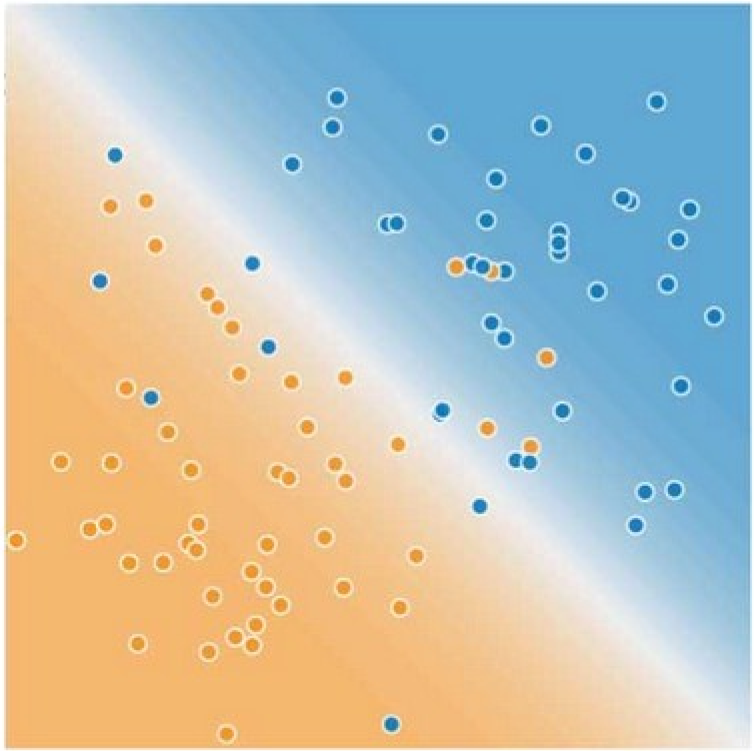
The datasets

- Training Dataset: The sample of data used to fit the model. The actual dataset that we use to train the model. The model sees and learns from this data.
- Test Dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

At actual...



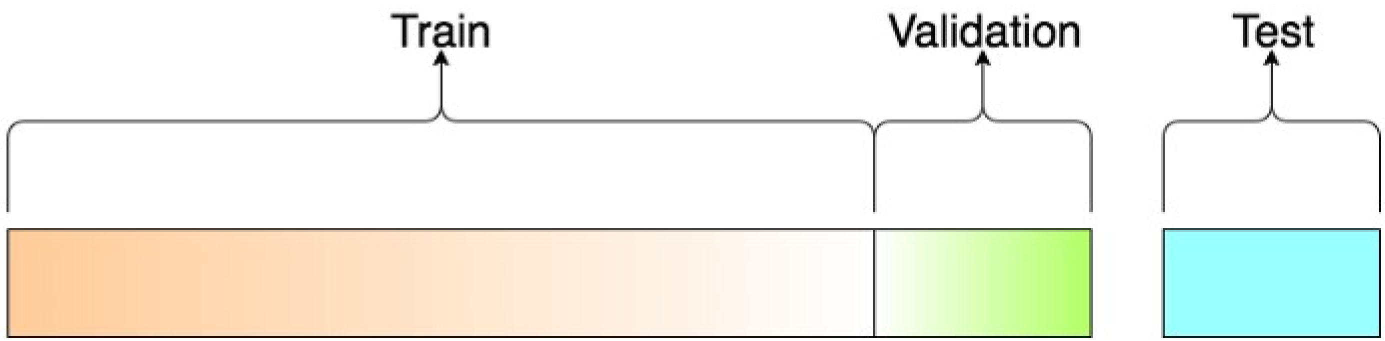
Training Data



Test Data

Validation Dataset

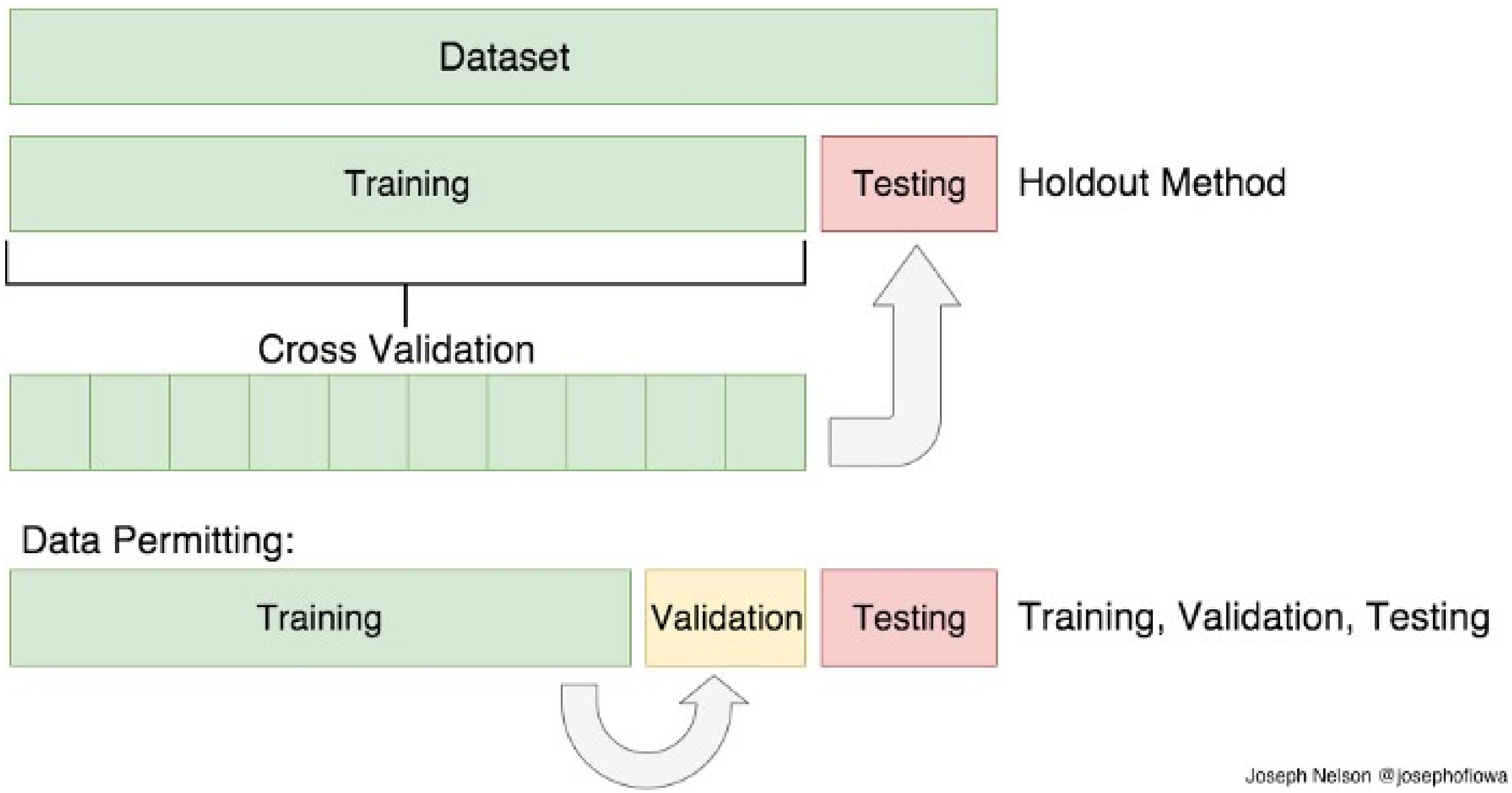
- Validation Dataset: The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters.
- The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.



Validation Dataset

- The validation set is used to evaluate a given model, but this is for frequent evaluation.
- We, as machine learning engineers, use this data to fine-tune the model hyperparameters. Hence the model occasionally sees this data, but never does it “Learn” from this.
- We use the validation set results, and update higher level hyperparameters. So the validation set affects a model, but only indirectly.
- The validation set is also known as the Dev set or the Development set. This makes sense since this dataset helps during the “development” stage of the model.

In short...



Dataset
Training
Testing
Holdout Method
Cross Validation
Data Permitting:
Training
Validation
Testing
Training, Validation, Testing
Joseph Nelson @josephofiowa

Actual situation

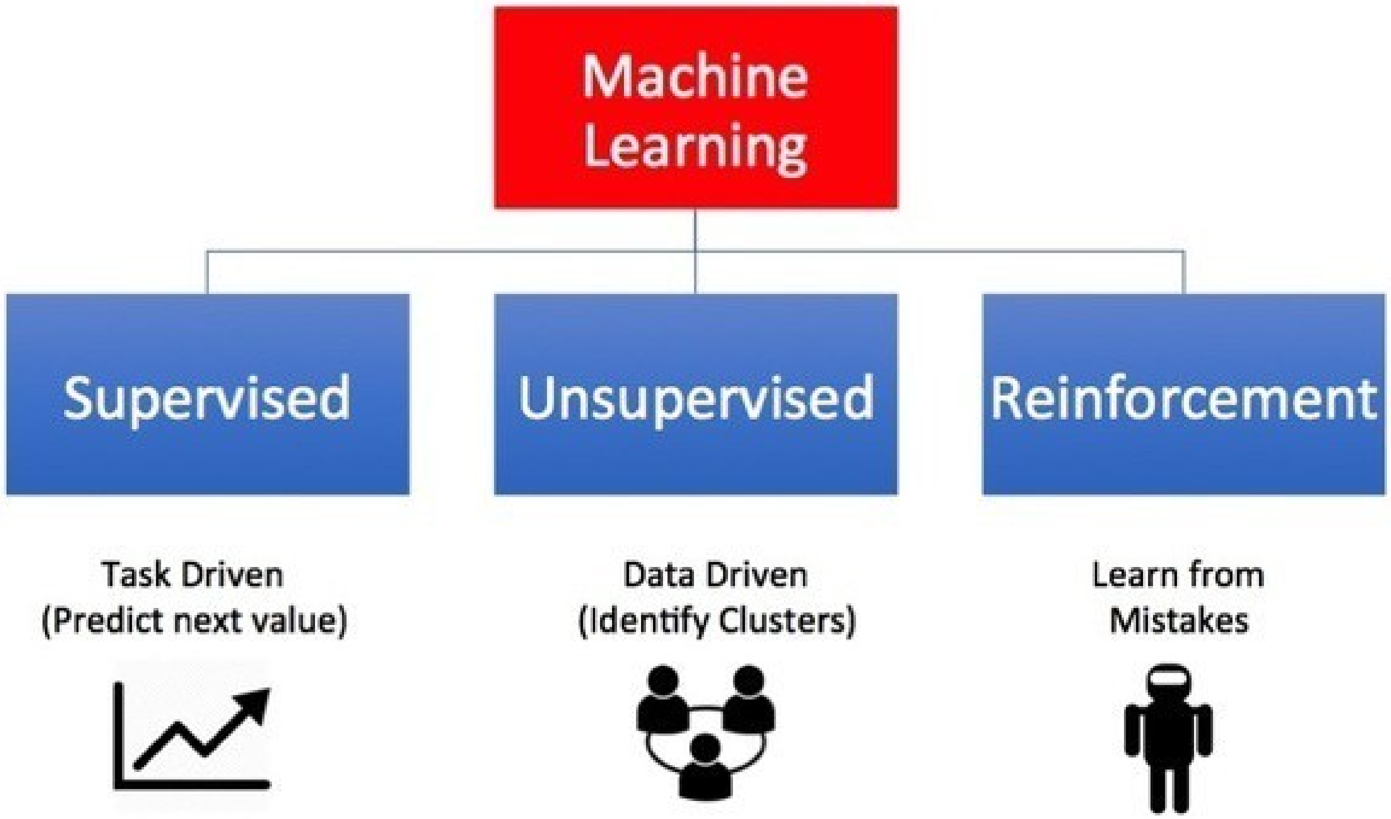
Subject	*t*	**Feature 1**	**Feature 2**	**Target**
Paul	1	1000	male	0
Paul	2	1100	male	0
Paul	3	1200	male	1
Paul	4	1300	male	1
Crista	4	20	female	0
Crista	5	100	female	0
Paulina	1	10000	female	0
Paulina	2	100000	female	1
Paulina	3	95000	female	1
Paulina	4	97000	female	1
Paulina	5	99000	female	1
Paulina	6	101000	female	1
George	1	50000	male	1
George	2	50000	male	1
George	3	50000	male	1
George	4	50000	male	1
George	5	50000	male	1
George	6	50000	male	1

Train

Test

Subject	*t*	**Feature 1**	**Feature 2**	**Target**
Paul	1	1000	male	0
Paulina	1	10000	female	0
George	1	50000	male	1
Paul	2	1100	male	0
Paulina	2	100000	female	1
George	2	50000	male	1
Paul	3	1200	male	1
Paulina	3	95000	female	1
George	3	50000	male	1
Paul	4	1300	male	1
Crista	4	20	female	0
Paulina	4	97000	female	1
George	4	50000	male	1
Crista	5	100	female	0
Paulina	5	99000	female	1
George	5	50000	male	1
Paulina	6	101000	female	1
George	6	50000	male	1

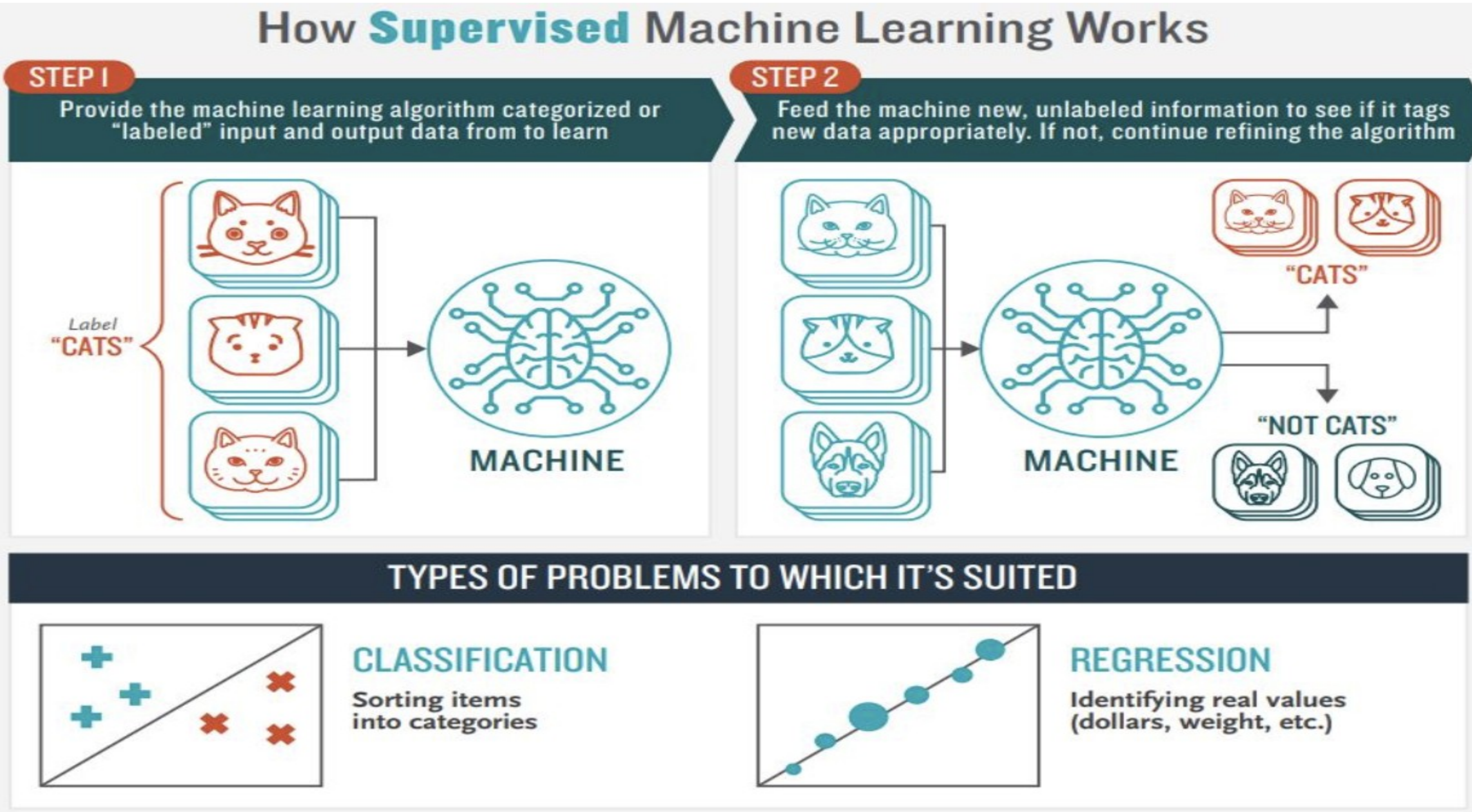
Types of Machine Learning



Supervised Learning

- Supervised learning (SL) is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.
- It infers a function from labeled training data consisting of a set of training examples.
- In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal).
- A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

Supervised Machine Learning



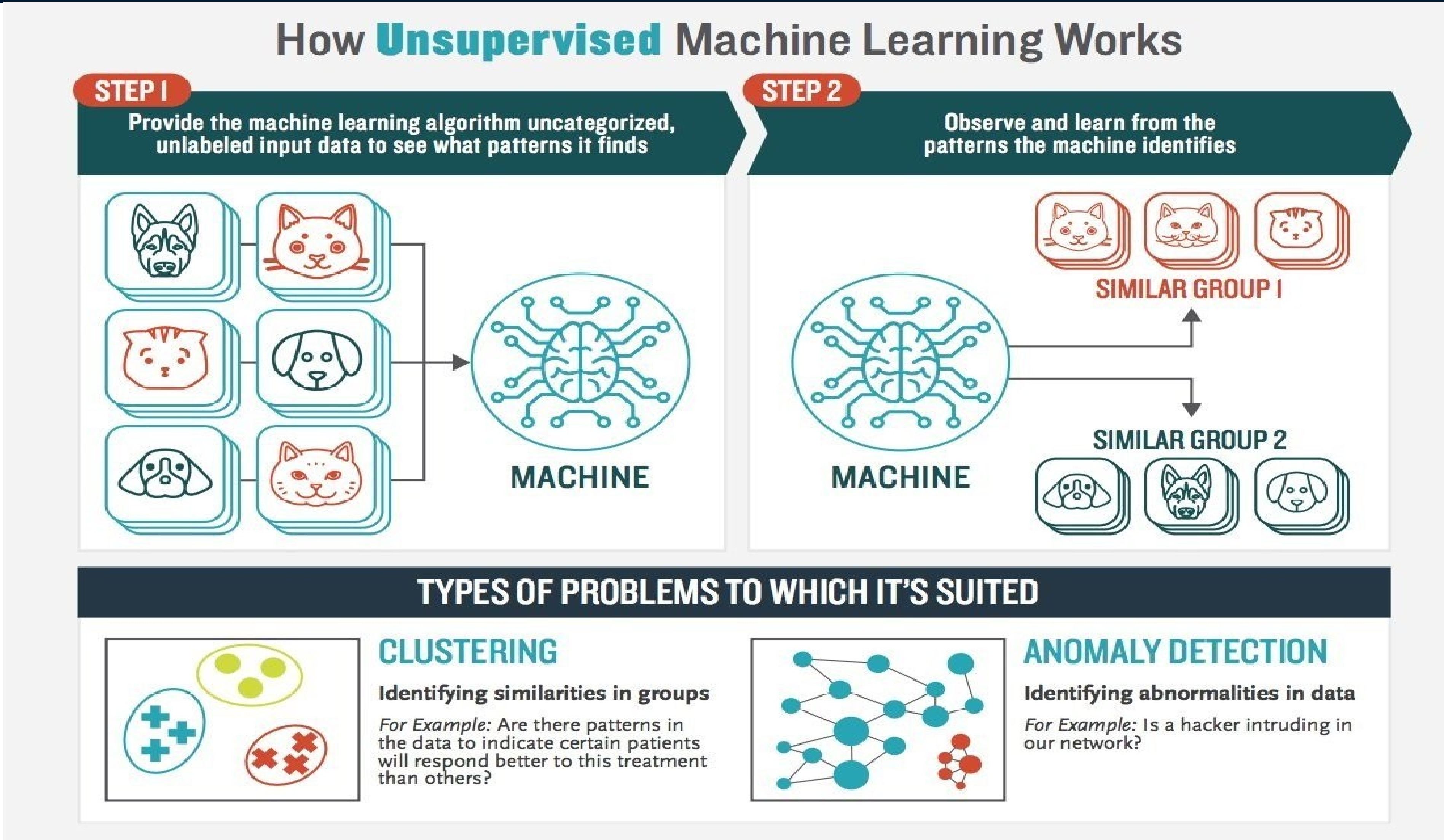
Supervised Learning : Examples

- Support-vector machines
- Linear regression
- Logistic regression
- Naive Bayes
- Linear discriminant analysis
- Decision trees
- K-nearest neighbor algorithm
- Neural networks (Multilayer perceptron)

Un-Supervised Learning

- Unsupervised learning (UL) is a type of algorithm that learns patterns from untagged data.
- The hope is that, through mimicry, the machine is forced to build a compact internal representation of its world and then generate imaginative content.
- In contrast to supervised learning (SL) where data is tagged by a human, e.g. as "car" or "fish" etc, UL exhibits self-organization that captures patterns as neuronal predilections or probability densities.

Unsupervised Machine Learning



Un-Supervised Learning : Examples

- Clustering methods include: hierarchical clustering, k-means, mixture models, DBSCAN, and OPTICS algorithm
- Anomaly detection methods include: Local Outlier Factor, and Isolation Forest
- Learning latent variable models such as Expectation–maximization algorithm (EM), Method of moments, and Blind signal separation techniques (Principal component analysis, Independent component analysis, Non-negative matrix factorization, Singular value decomposition)

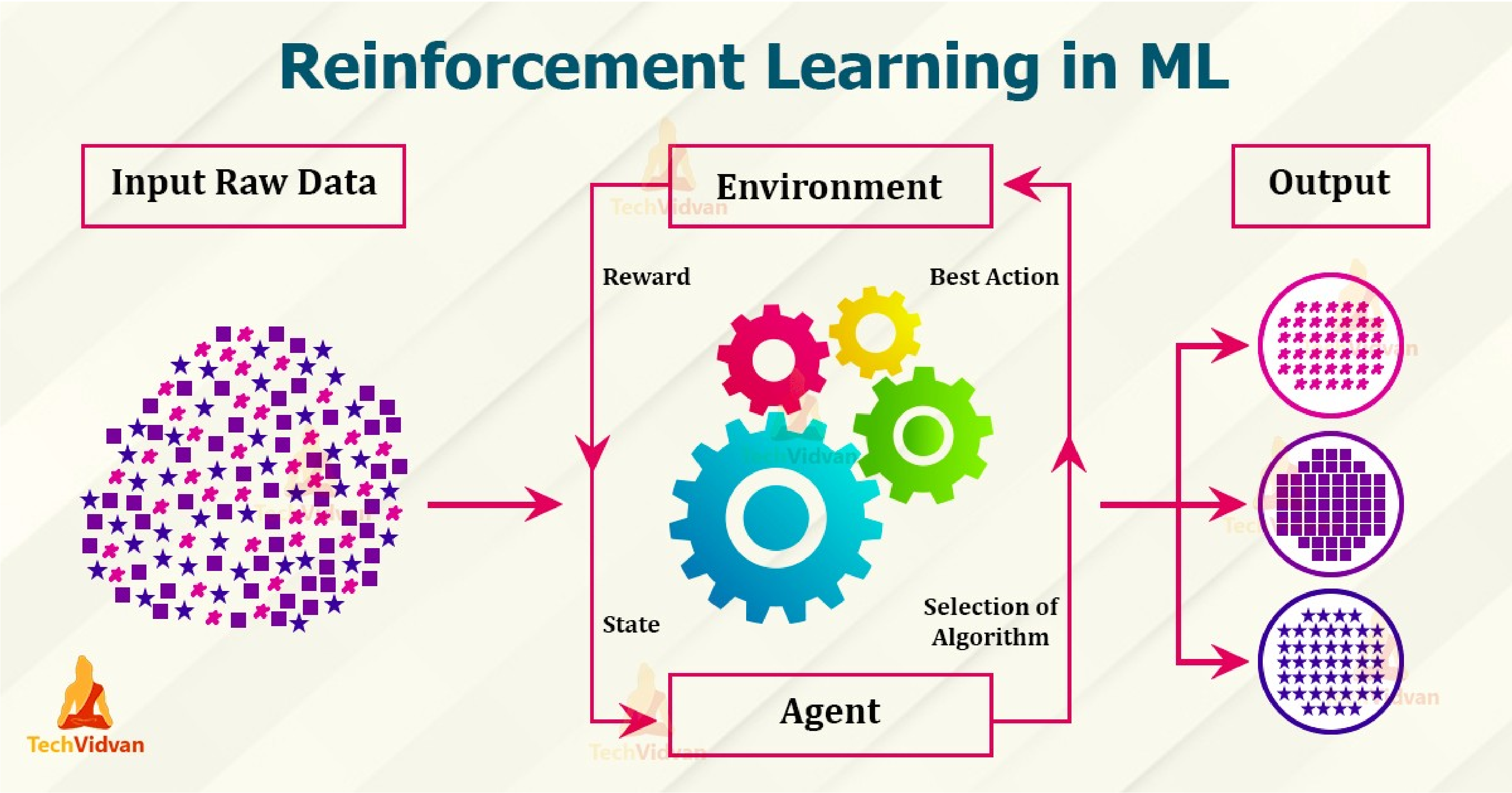
Reinforcement Learning

- Reinforcement Learning is defined as a Machine Learning method that is concerned with how software agents should take actions in an environment.
- Reinforcement Learning is a part of the deep learning method that helps you to maximize some portion of the cumulative reward.

Reinforcement Learning

- Imagine someone playing a video game. The player is the agent, and the game is the environment. The rewards the player gets (i.e. beat an enemy, complete a level), or doesn't get (i.e. step into a trap, lose a fight) will teach him how to be a better player.
- In supervised learning, for example, each decision taken by the model is independent, and doesn't affect what we see in the future.
- In reinforcement learning, instead, we are interested in a long term strategy for our agent, which might include sub-optimal decisions at intermediate steps, and a trade-off between exploration (of unknown paths), and exploitation of what we already know about the environment.

Reinforcement Learning



History

- For several decades (since the 1950s!), reinforcement learning followed two separate threads of research, one focusing on trial and error approaches, and one based on optimal control.
- Optimal control methods are aimed at designing a controller to minimize a measure of a dynamical system's behaviour over time.
- To achieve this, they mainly used dynamic programming algorithms, which we will see are the foundations of modern reinforcement learning techniques.

History

- Trial-and-error approaches, instead, have deep roots in the psychology of animal learning and neuroscience, and this is where the term reinforcement comes from: actions followed (reinforced) by good or bad outcomes have the tendency to be reselected accordingly.
- Arising from the interdisciplinary study of these two fields came a field called Temporal Difference (TD) Learning.
- The modern machine learning approaches to RL are mainly based on TD-Learning, which deals with rewards signals and a value function

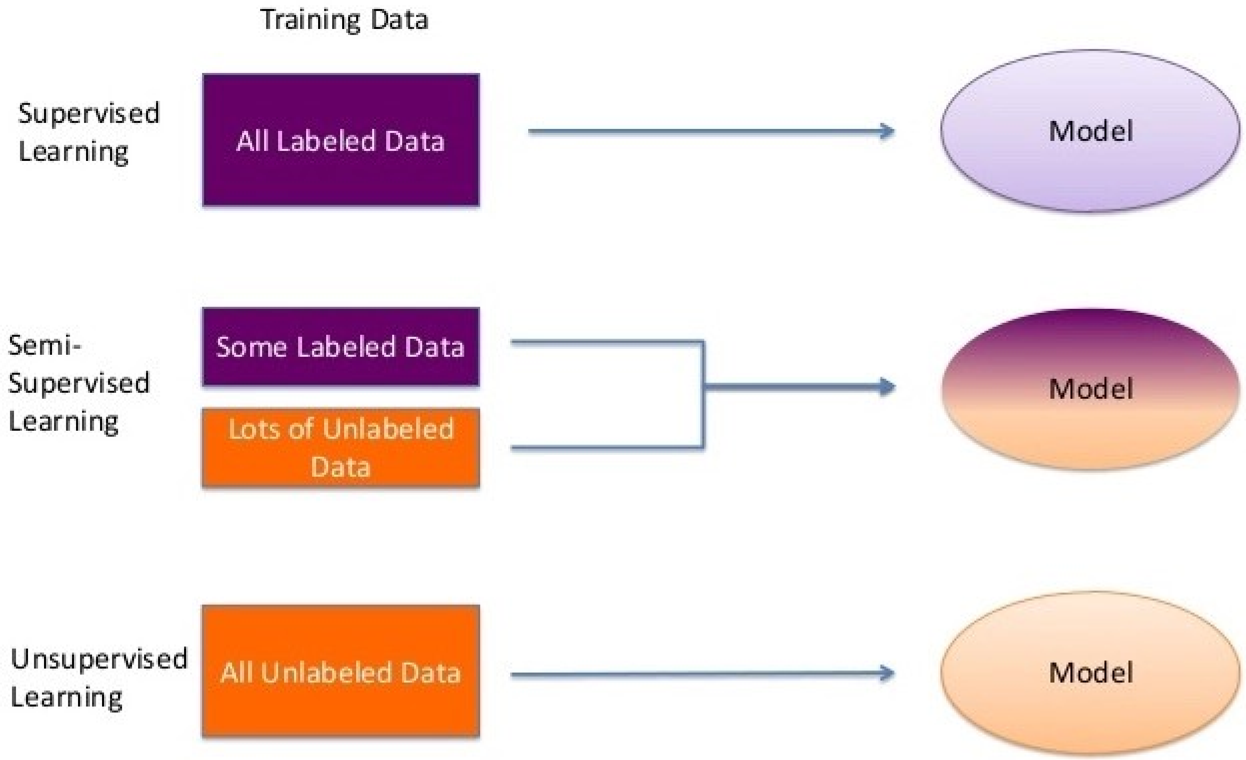
Further Categorization

- Hybrid Learning Problems
    4. Semi-Supervised Learning
    5. Self-Supervised Learning
    6. Multi-Instance Learning
- Statistical Inference
    7. Inductive Learning
    8. Deductive Inference
    9. Transductive Learning
- Learning Techniques
    10. Multi-Task Learning
    11. Active Learning
    12. Online Learning
    13. Transfer Learning
    14. Ensemble Learning

Semi Supervised Learning

- Semi-supervised learning is supervised learning where the training data contains very few labeled examples and a large number of unlabeled examples.
- The goal of a semi-supervised learning model is to make effective use of all of the available data, not just the labelled data like in supervised learning.
- In semi-supervised learning we are given a few labeled examples and must make what we can of a large collection of unlabeled examples. Even the labels themselves may not be the oracular truths that we hope for.

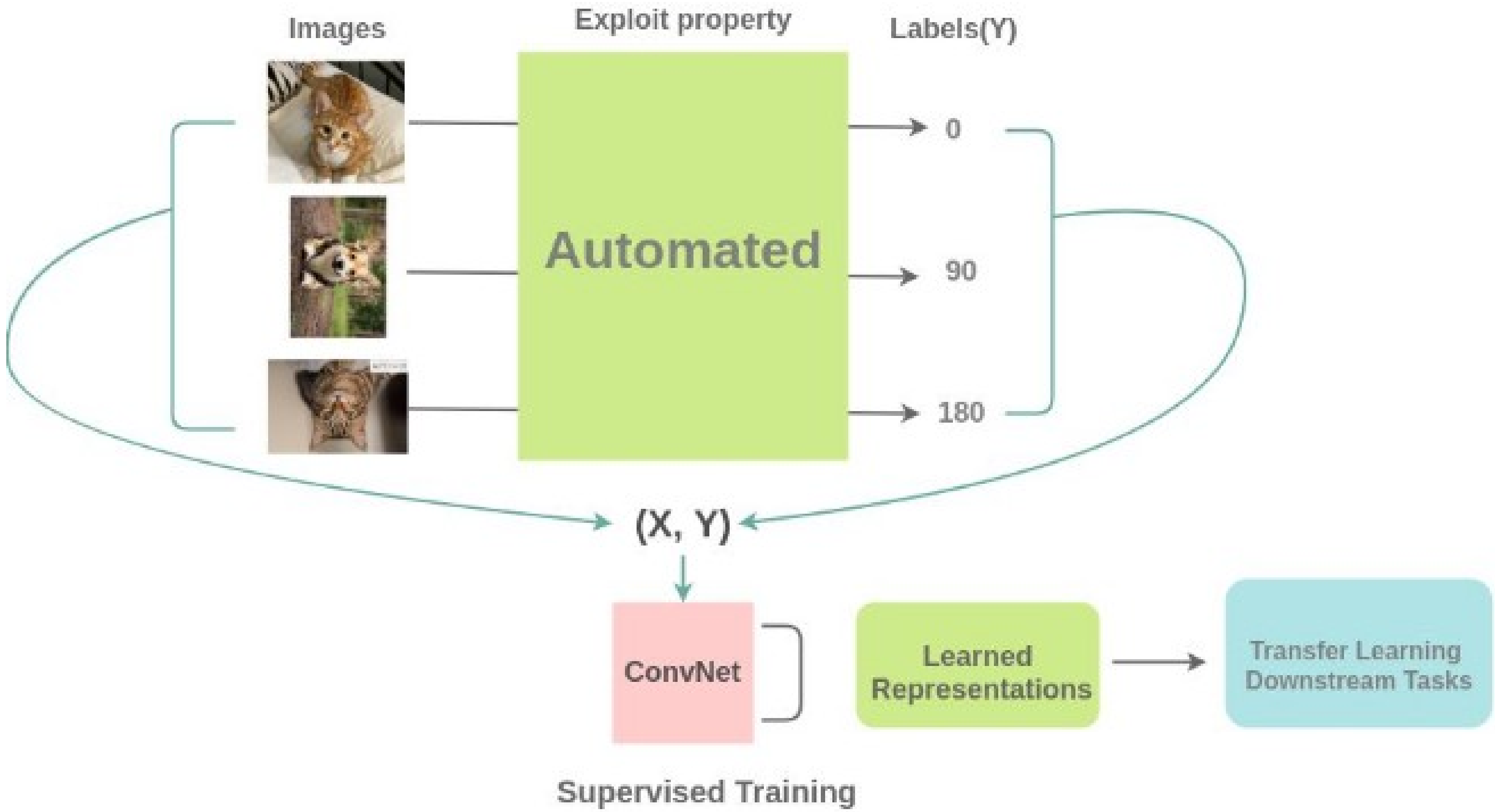
Semi Supervised Learning



Self Supervised Learning

- Self-supervised learning refers to an unsupervised learning problem that is framed as a supervised learning problem in order to apply supervised learning algorithms to solve it.
- Supervised learning algorithms are used to solve an alternate or pretext task, the result of which is a model or representation that can be used in the solution of the original (actual) modeling problem.
- The self-supervised learning framework requires only unlabeled data in order to formulate a pretext learning task such as predicting context or image rotation, for which a target objective can be computed without supervision.

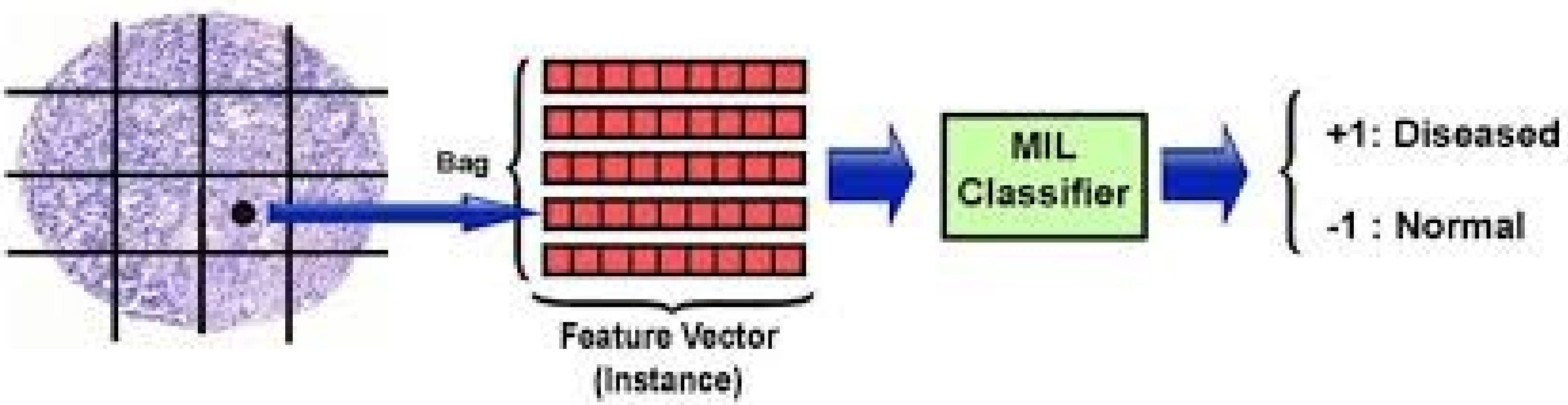
Self Supervised Learning



Multi Instance Learning

- Multi-instance learning is a supervised learning problem where individual examples are unlabeled; instead, bags or groups of samples are labeled.
- In multi-instance learning, an entire collection of examples is labeled as containing or not containing an example of a class, but the individual members of the collection are not labeled.

Multi Instance Learning



Stastical Inference

- Inference refers to reaching an outcome or decision.
- In machine learning, fitting a model and making a prediction are both types of inference.
- There are different paradigms for inference that may be used as a framework for understanding how some machine learning algorithms work or how some learning problems may be approached.

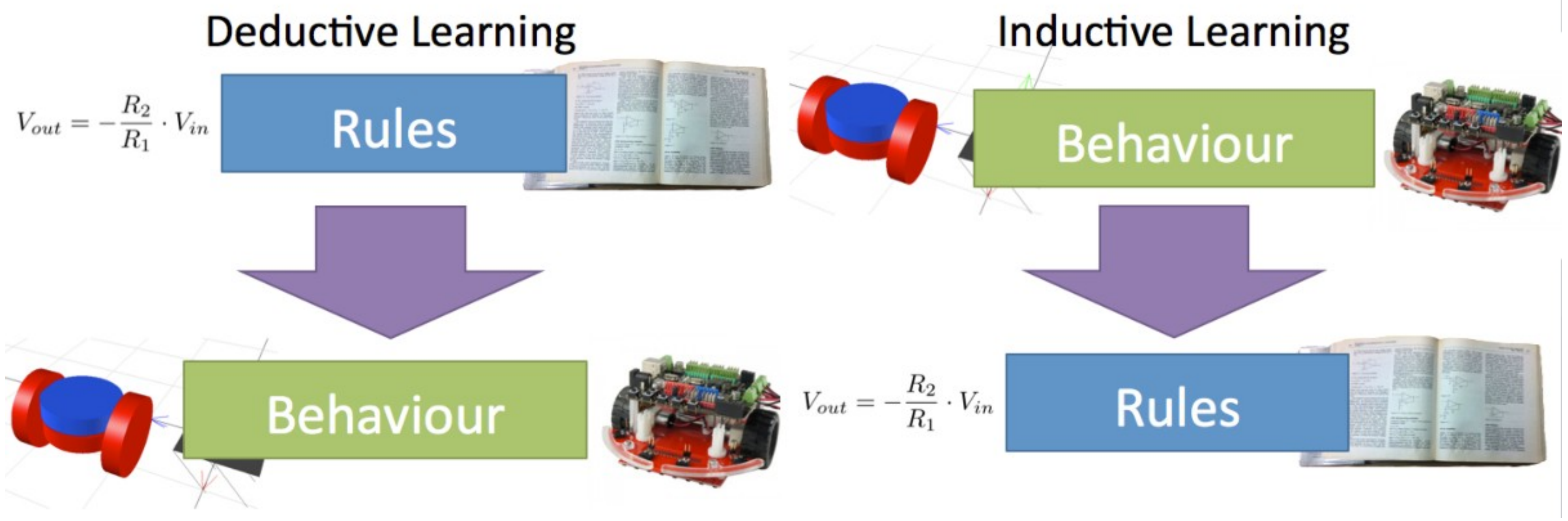
Inductive Learning

- Inductive learning involves using evidence to determine the outcome.
- Inductive reasoning refers to using specific cases to determine general outcomes, e.g. specific to general.
- Most machine learning models learn using a type of inductive inference or inductive reasoning where general rules (the model) are learned from specific historical examples (the data).
  - … the problem of induction, which is the problem of how to draw general conclusions about the future from specific observations from the past.

Deductive Inference

- Deduction or deductive inference refers to using general rules to determine specific outcomes.
- We can better understand induction by contrasting it with deduction.
- Deduction is the reverse of induction. If induction is going from the specific to the general, deduction is going from the general to the specific.
  - … the simple observation that induction is just the inverse of deduction!

Comparing both



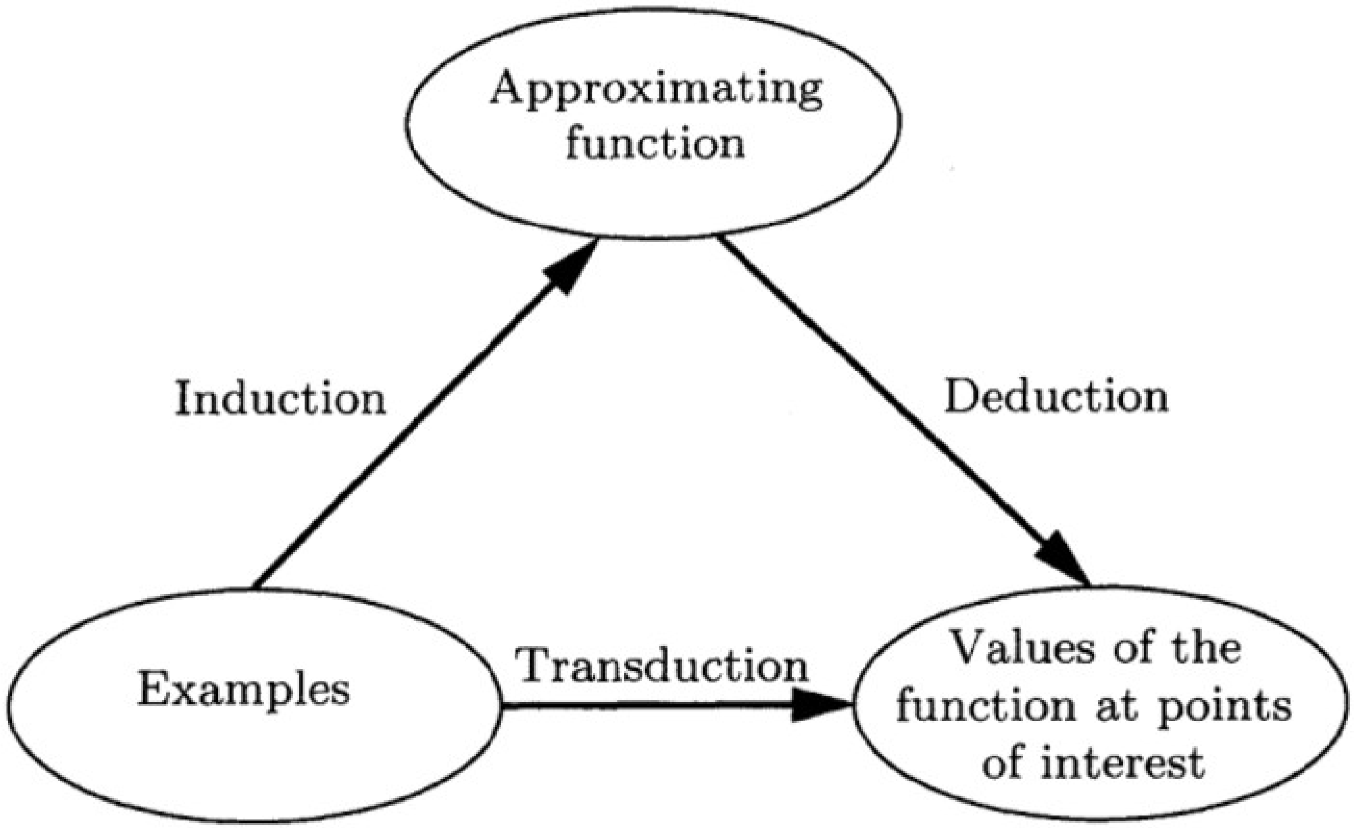
Transductive Learning

- Transduction or transductive learning is used in the field of statistical learning theory to refer to predicting specific examples given specific examples from a domain.
- It is different from induction that involves learning general rules from specific examples, e.g. specific to specific.
- Induction, deriving the function from the given data. Deduction, deriving the values of the given function for points of interest. Transduction, deriving the values of the unknown function for points of interest from the given data.

Induction, Deduction and Transduction

- We can contrast these three types of inference in the context of machine learning.
- For example:
  - Induction: Learning a general model from specific examples.
  - Deduction: Using a model to make predictions.
  - Transduction: Using specific examples to make predictions.

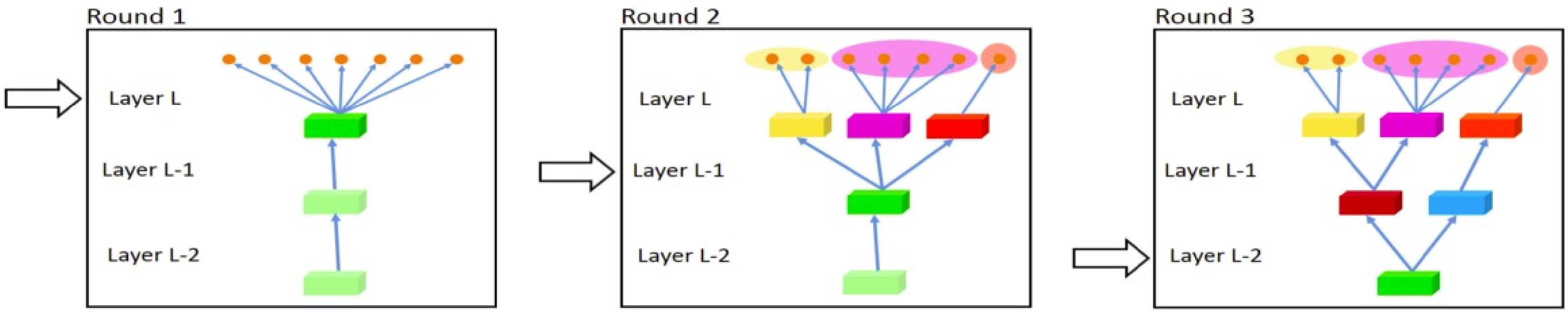
Induction, Deduction and Transduction



Multi-Task Learning

- Multi-task learning is a type of supervised learning that involves fitting a model on one dataset that addresses multiple related problems.
- It involves devising a model that can be trained on multiple related tasks in such a way that the performance of the model is improved by training across the tasks as compared to being trained on any single task.
- Multi-task learning is a way to improve generalization by pooling the examples (which can be seen as soft constraints imposed on the parameters) arising out of several tasks.

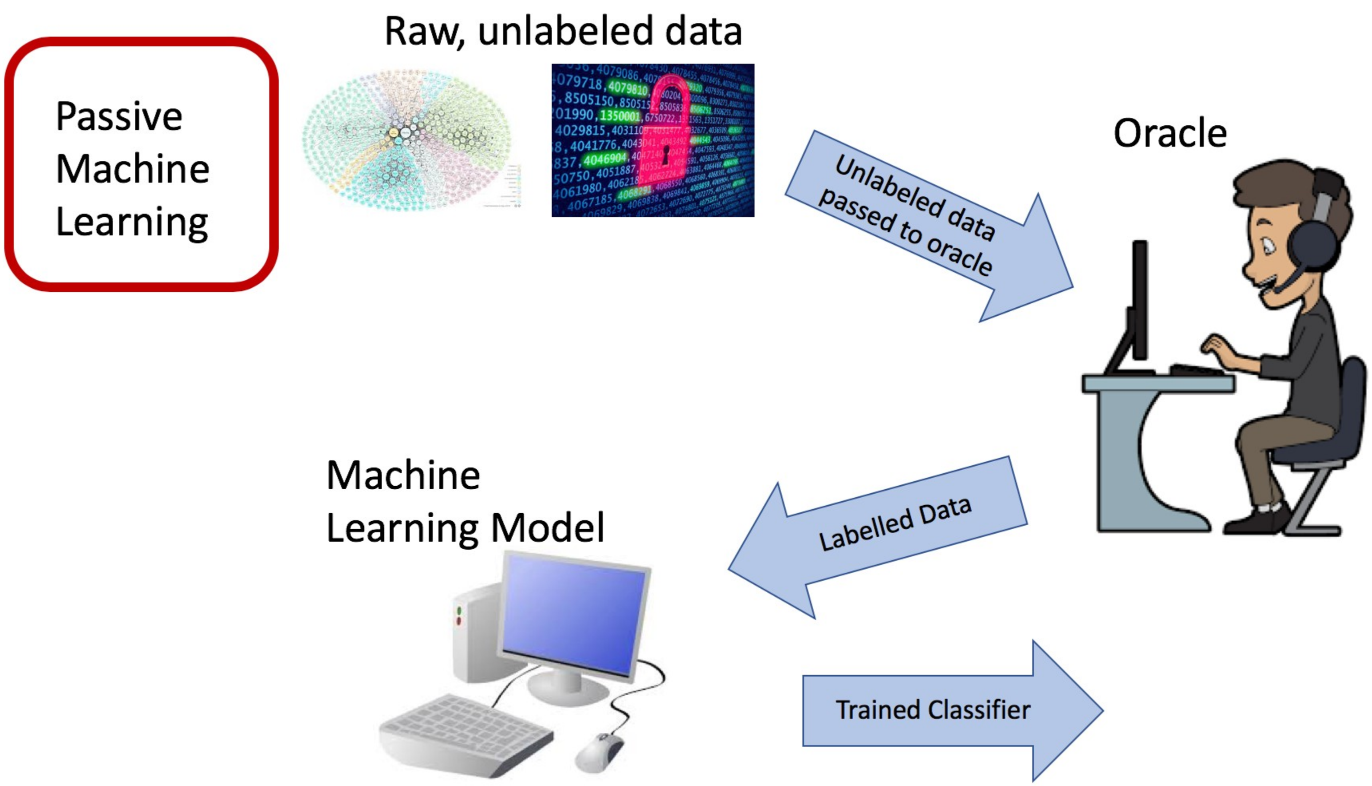
Multi-Task Learning



Active Learning

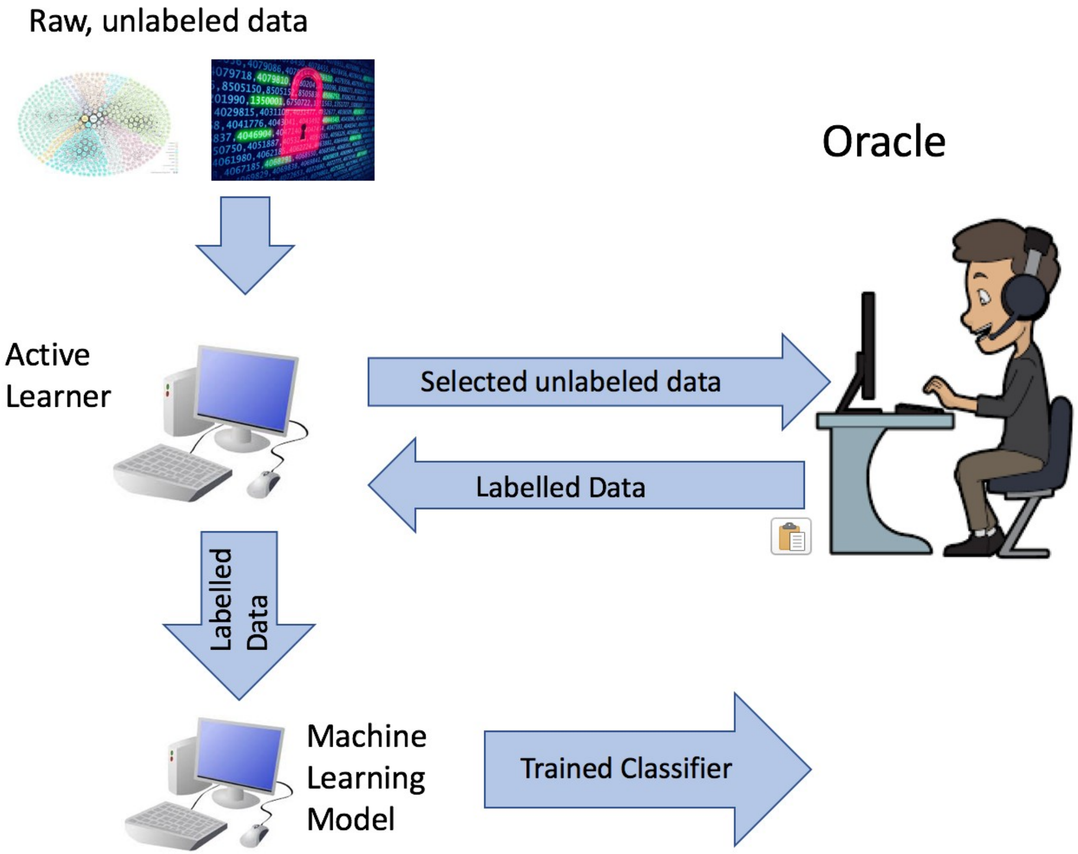
- Active learning is a technique where the model is able to query a human user operator during the learning process in order to resolve ambiguity during the learning process.
- Active learning: The learner adaptively or interactively collects training examples, typically by querying an oracle to request labels for new points.
- Active learning is a type of supervised learning and seeks to achieve the same or better performance of so-called “passive” supervised learning, although by being more efficient about what data is collected or used by the model.

Passive Learning



Active Learning

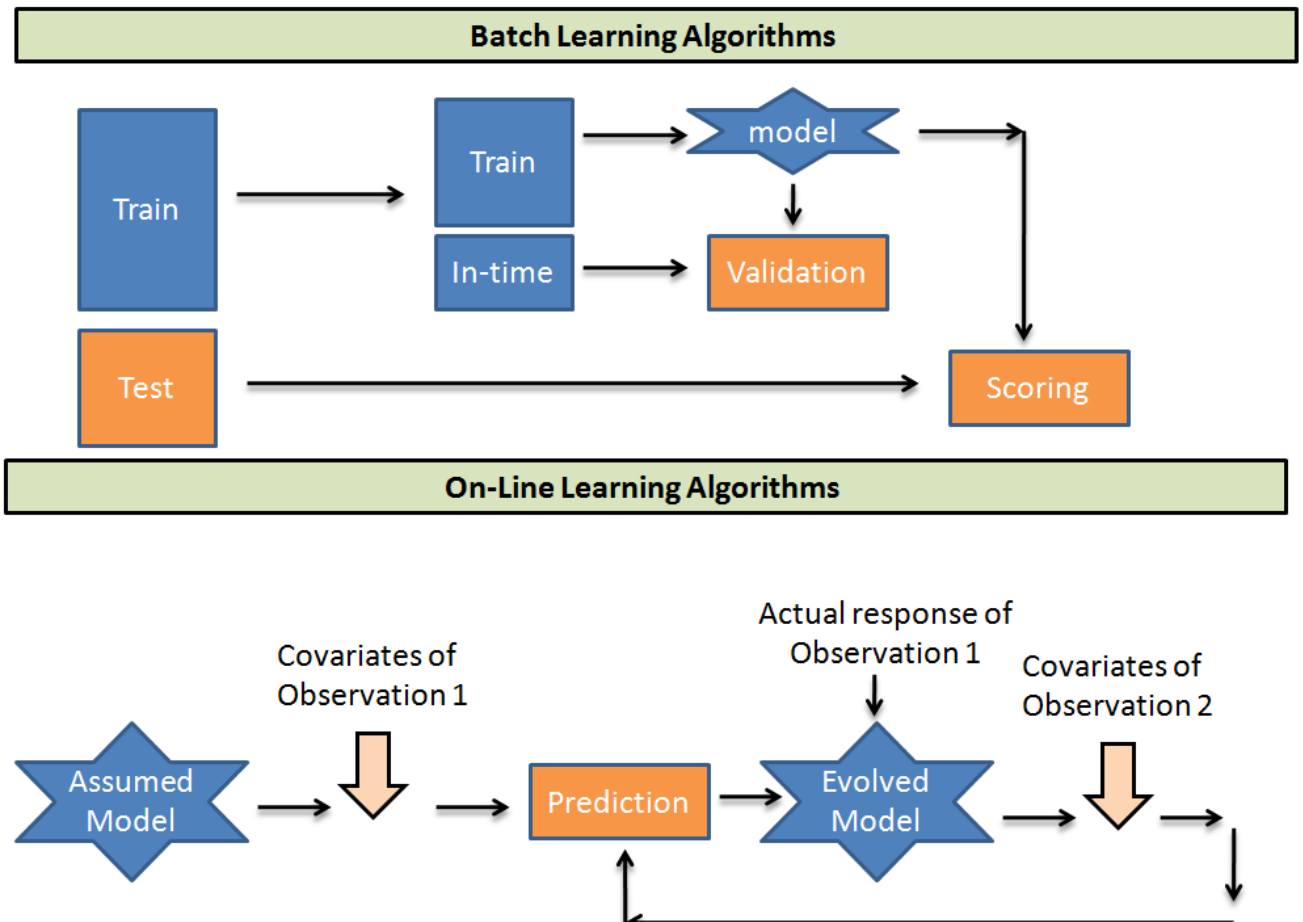
Active Machine Learning



Online Learning

- Online learning involves using the data available and updating the model directly before a prediction is required or after the last observation was made.
- Online learning is appropriate for those problems where observations are provided over time and where the probability distribution of observations is expected to also change over time.
- Therefore, the model is expected to change just as frequently in order to capture and harness those changes.

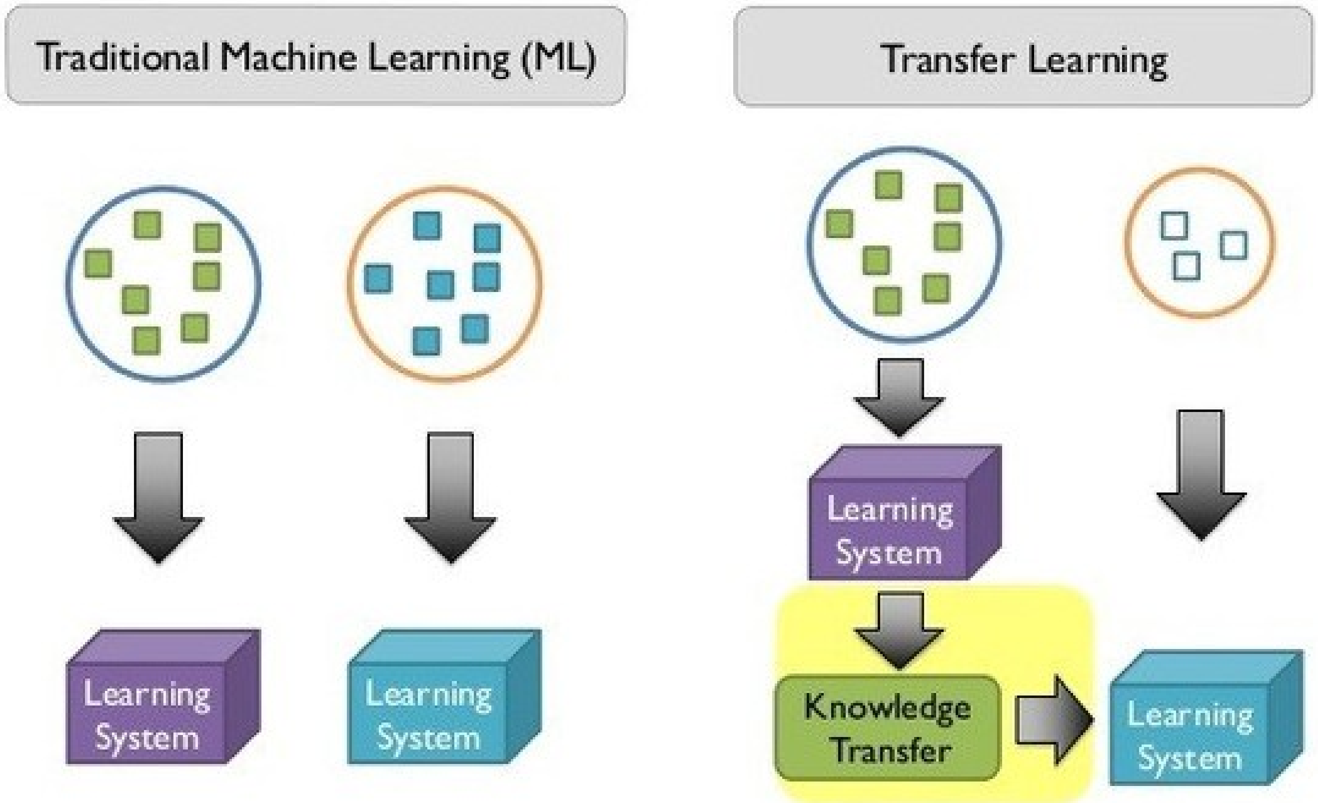
Online Learning



Transfer Learning

- Transfer learning is a type of learning where a model is first trained on one task, then some or all of the model is used as the starting point for a related task.
- In transfer learning, the learner must perform two or more different tasks, but we assume that many of the factors that explain the variations in P1 are relevant to the variations that need to be captured for learning P2.

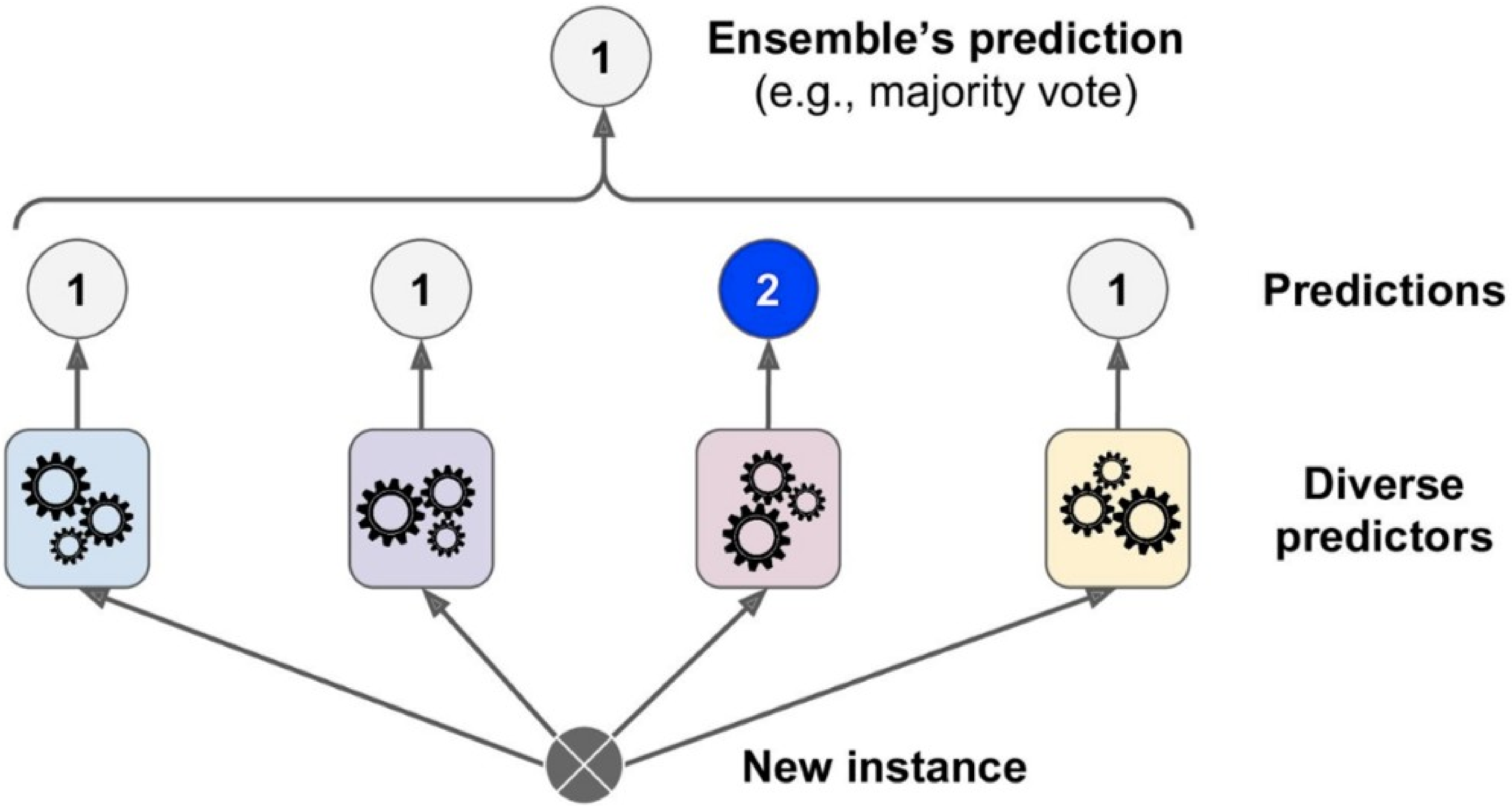
Transfer Learning



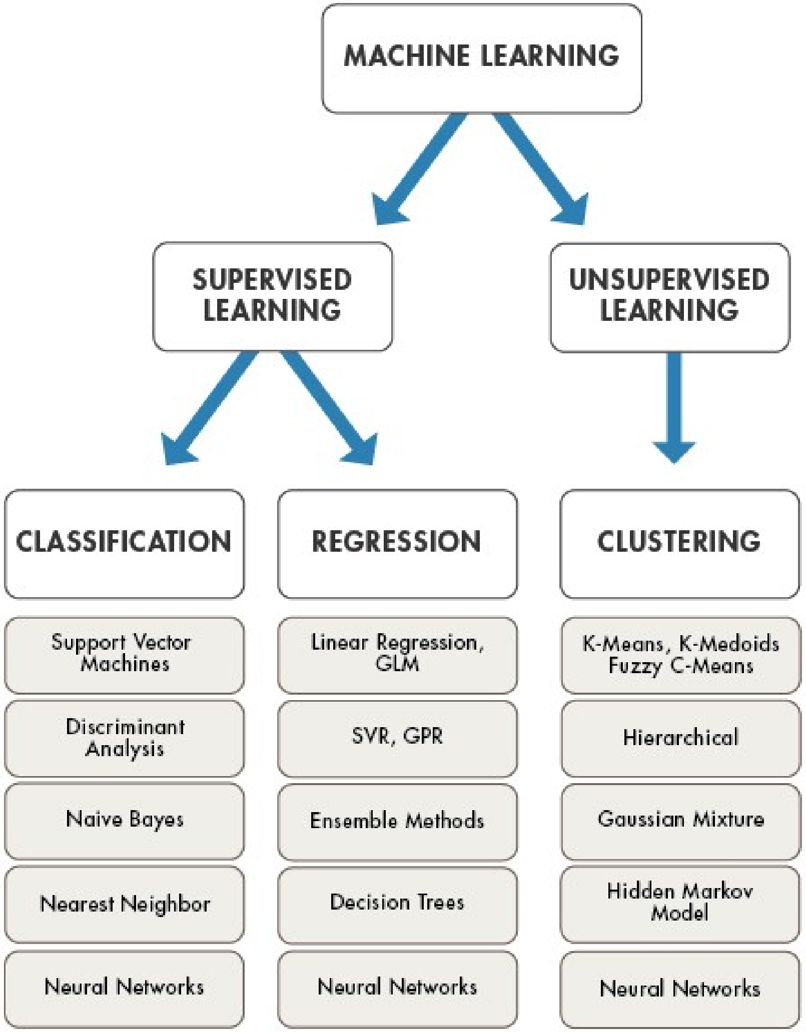
Ensemble Learning

- Ensemble learning is an approach where two or more modes are fit on the same data and the predictions from each model are combined.
- The field of ensemble learning provides many ways of combining the ensemble members’ predictions, including uniform weighting and weights chosen on a validation set.

Ensemble Learning



Typical Algorithms in Machine Learning



Predictive Analytics

- Predictive Analytics will help an organization to know what might happen next, it predicts future based on present data available.
- It will analyze the data and provide statements that have not happened yet.
- It makes all kinds of predictions that you want to know and all predictions are probabilistic in nature.

Descriptive Analytics

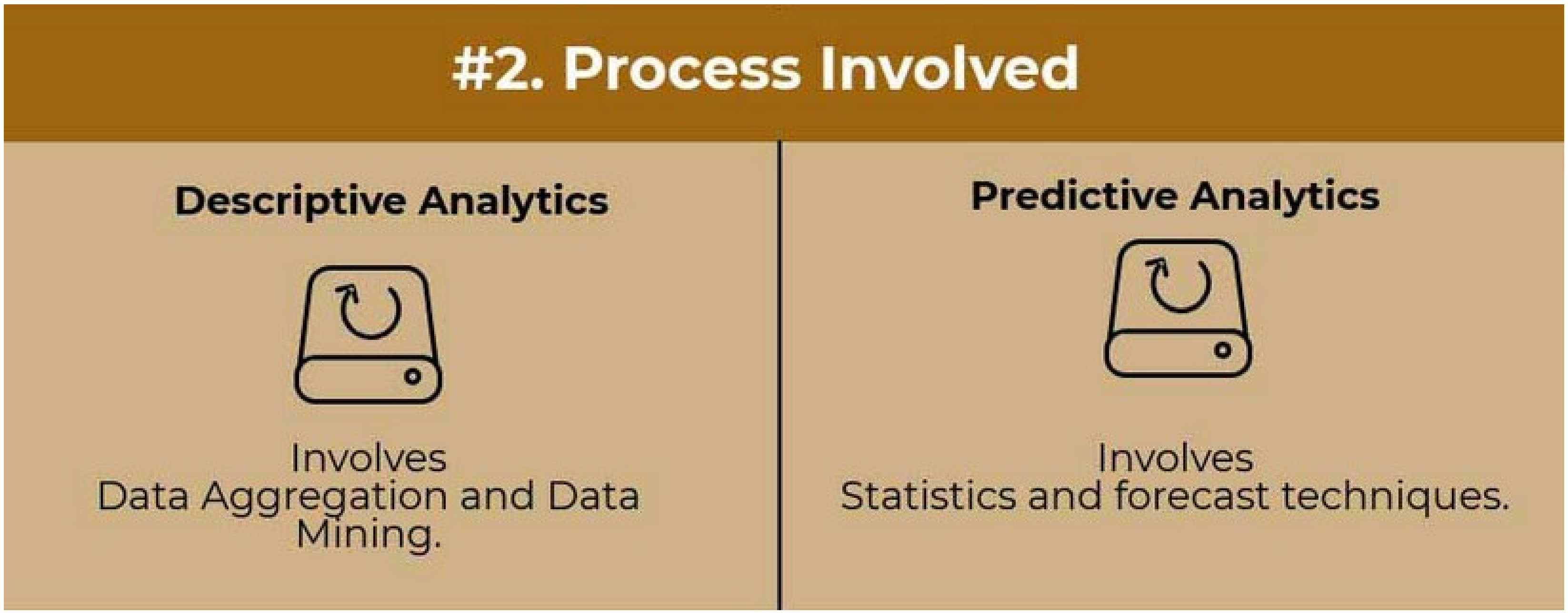
- Descriptive Analytics will help an organization to know what has happened in the past, it would give you the past analytics using the data that are stored.
- For a company, it is necessary to know the past events that help them to make decisions based on the statistics using historical data.
- For example, you might want to know how much money you lost due to fraud and many more.

Descriptive vs. Predictive Analytics

1#. Describes

Descriptive Analytics	Predictive Analytics
What happened in the past? By using the stored data.	What might happen in the future? By using the past data and analysing it.

Descriptive vs. Predictive Analytics

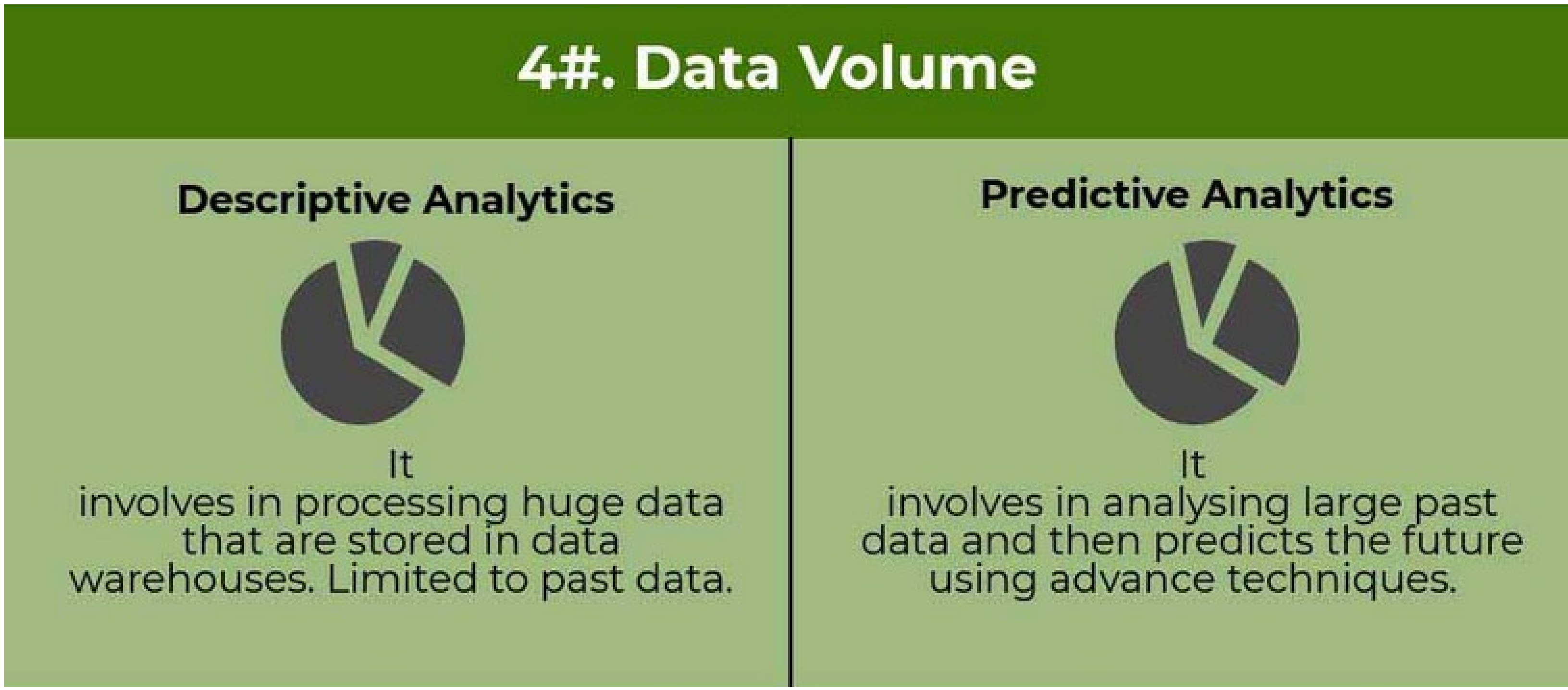


Descriptive vs. Predictive Analytics

3#. Definition

Descriptive Analytics	Predictive Analytics
The process of finding the useful and important information by analysing the huge data.	This process involves in forecasting the future of the company, which are very useful.

Descriptive vs. Predictive Analytics

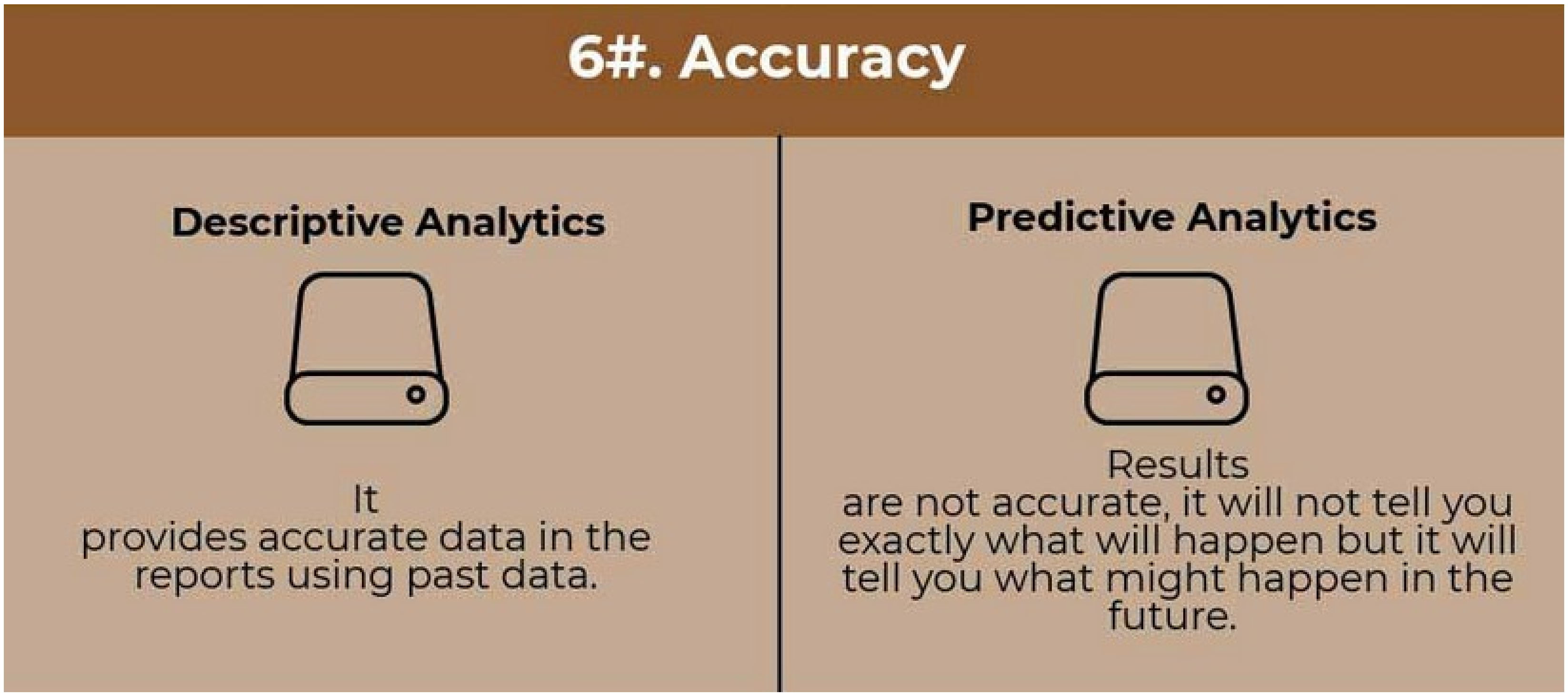


Descriptive vs. Predictive Analytics

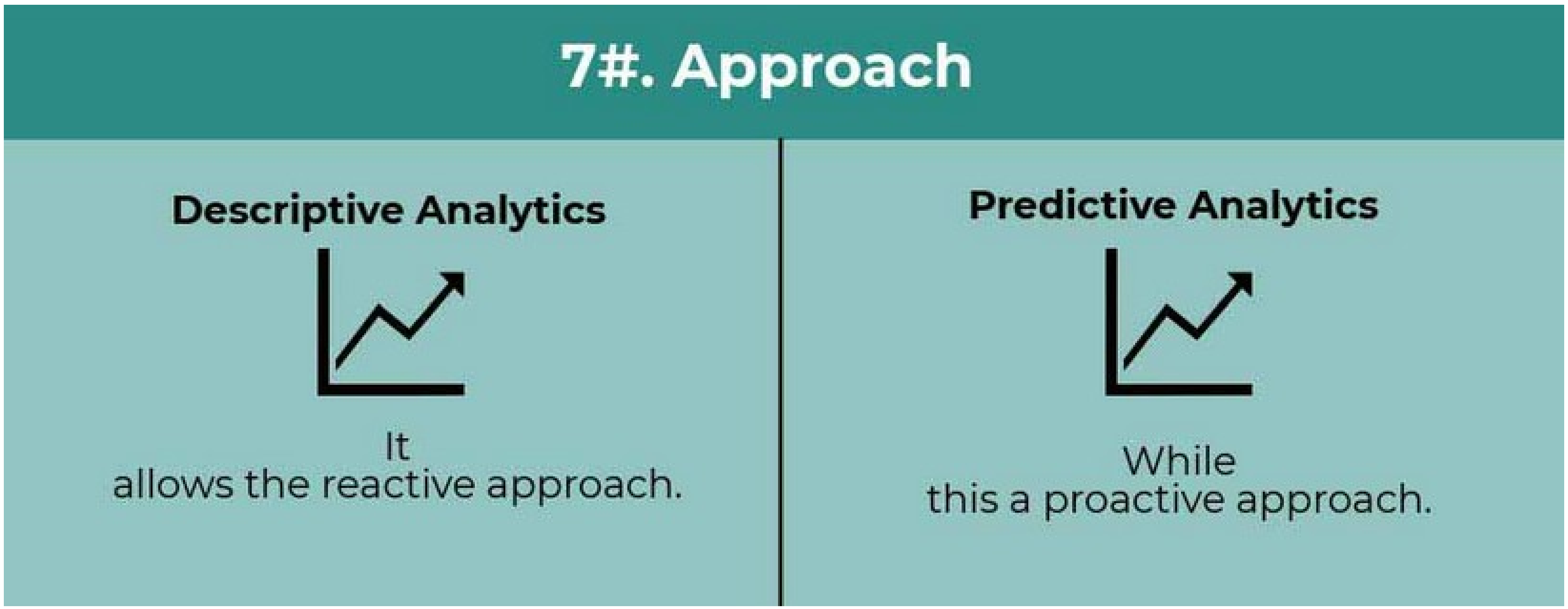
5#. Examples

Descriptive Analytics	Predictive Analytics
Sales report, revenue of a company, performance analysis, etc.	Sentimental analysis, credit score analysis, forecast reports for company, etc.

Descriptive vs. Predictive Analytics



Descriptive vs. Predictive Analytics



Summary

- Descriptive analysis is centered around the presentation of data, visualization to the management sights. While predictive analysis is centered around statistical model which helps to predict the future.
- The predictive analysis has more risk as it involves in analyzing what exactly will happen in the future based on the past events, but the certain condition may not happen exactly in the future for the same reason.

Designing a Learning System

- Choosing the Training Experience
- Choosing the Target Function
- Choosing a Representation for the Target Function
- Choosing a Function Approximation Algorithm
- The Final Design

Choosing the Training Experience

- Whether the training experience provides direct or indirect feedback regarding the choices made by the performance system:
- Example :
  - Direct training examples in learning to play checkers consist of individual checkers board states and the correct move for each.
  - Indirect training examples in the same game consist of the move sequences and final outcomes of various games played in which information about the correctness of specific moves early in the game must be inferred indirectly from the fact that the game was eventually won or lost – credit assignment problem.

Choosing the Training Experience

- The degree to which the learner controls the sequence of training examples:
- Example :
  - The learner might rely on the teacher to select informative board states and to provide the correct move for each
  - The learner might itself propose board states that it finds particularly confusing and ask the teacher for the correct move. Or the learner may have complete control over the board states and (indirect) classifications, as it does when it learns by playing against itself with no teacher present.

Choosing the Training Experience

- How well it represents the distribution of examples over which the final system performance P must be measured: In general learning is most reliable when the training examples follow a distribution similar to that of future test examples.
- Example :
  - If the training experience in play checkers consists only of games played against itself, the learner might never encounter certain crucial board states that are very likely to be played by the human checkers champion. (Note however that the most current theory of machine learning rests on the crucial assumption that the distribution of training examples is identical to the distribution of test examples)

Choosing the Target Function

- To determine what type of knowledge will be learned and how this will be used by the performance program:
- Example :
  - In play checkers, it needs to learn to choose the best move among those legal moves: ChooseMove: B → M, which accepts as input any board from the set of legal board states B and produces as output some move from the set of legal moves M.

Choosing the Target Function

- Since the target function such as ChooseMove turns out to be very difficult to learn given the kind of indirect training experience available to the system, an alternative target function is then an evaluation function that assigns a numerical score to any given board state, $V: B \rightarrow R$.

Choosing a Representation for the Target Function

- Given the ideal target function V, we choose a representation that the learning system will use to describe V' that it will learn:
- Example :
  - In play checkers,

    $V'(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$
  - where w i is the numerical coefficient or weight to determine the relative importance of the various board features and x_i is the number of i-th objects on the board.

Choosing a Function Approximation Algorithm

- Each training example is given by <b, V train (b)> where V train (b) is the training value for a board b.
- Estimating Training Values:

$$V_{train}(b) \leftarrow V'(Successor(b)).$$

- Adjusting the weights: To specify the learning algorithm for choosing the weights w i to best fit the set of training examples {<b, V train (b)>}, which minimizes the squared error E between the training values and the values predicted by the hypothesis V‘

$$E = \sum_{\langle b, V_{train}(b)\rangle \in \text{training examples}} (V_{train}(b) - V'(b))^2$$

Choosing a Function Approximation Algorithm

$$E = \sum_{\langle b, V_{train}(b)\rangle \in \text{training examples}} (V_{train}(b) - V'(b))^2$$

- To minimize E, the following rule is used:

  LMS weight update rule

  For each training example <b, V train (b)> Use the current weights to calculate V'(b) For each weight w_i , update it as,

$$w_i \leftarrow w_i + \eta\,(V_{train}(b) - V'(b))\,x_i$$

The Final Design

- Performance System: To solve the given performance task by using the learned target function(s). It takes an instance of a new problem (new game) as input and a trace of its solution (game history) as output.
- Critic: To take as input the history or trace of the game and produce as output a set of training examples of the target function.

The Final Design

- Generalizer: To take as input the training examples and produce an output hypothesis that is its estimate of the target function. It generalizes from the specific training examples, hypothesizing a general function that covers these examples and other cases beyond the training examples.
- Experiment Generator: To take as input the current hypothesis (currently learned function) and outputs a new problem (i.e., initial board state) for Performance System to explore. Its role is to pick new practice problems that will maximize the learning rate of the overall system.

The Final Design

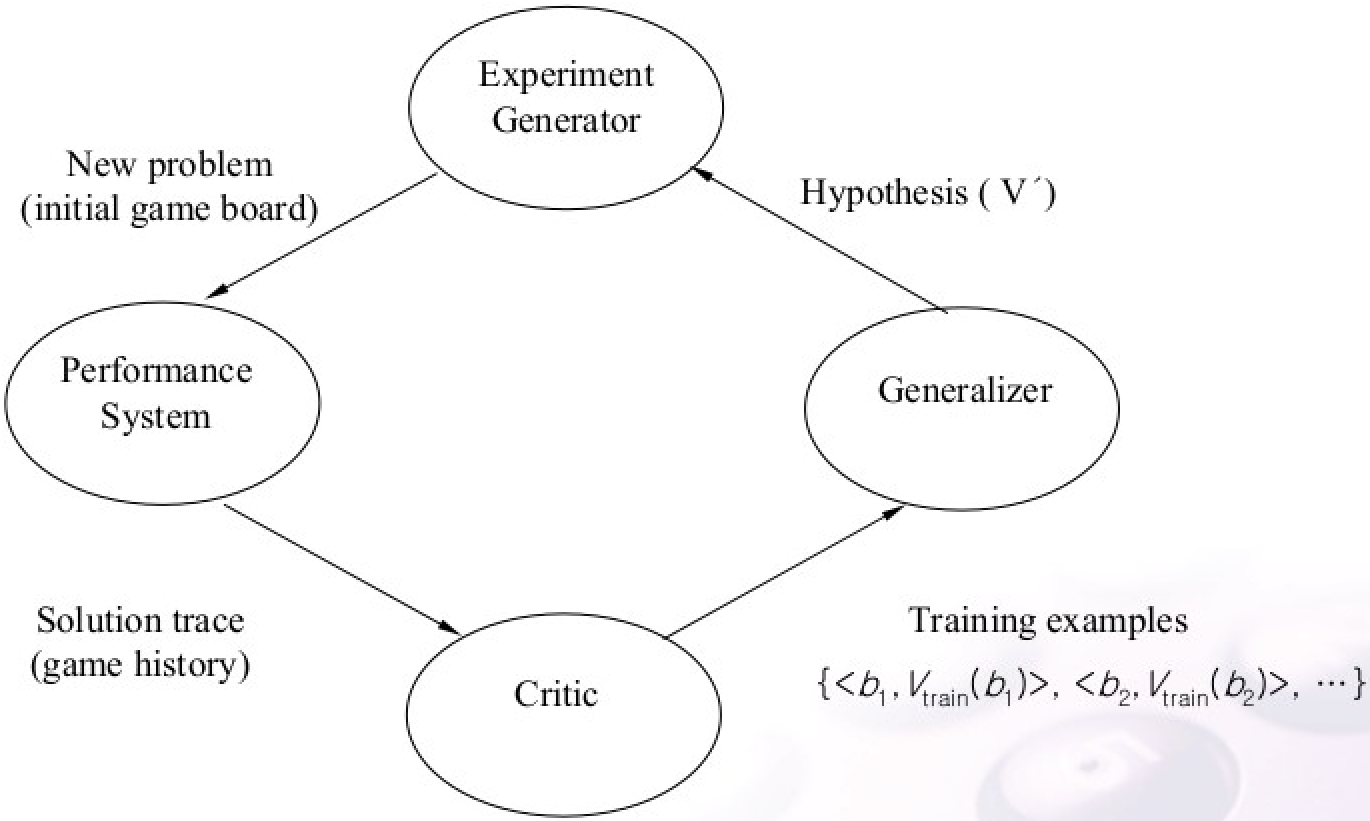


Figure 1.1 Final design of the checkers learning program

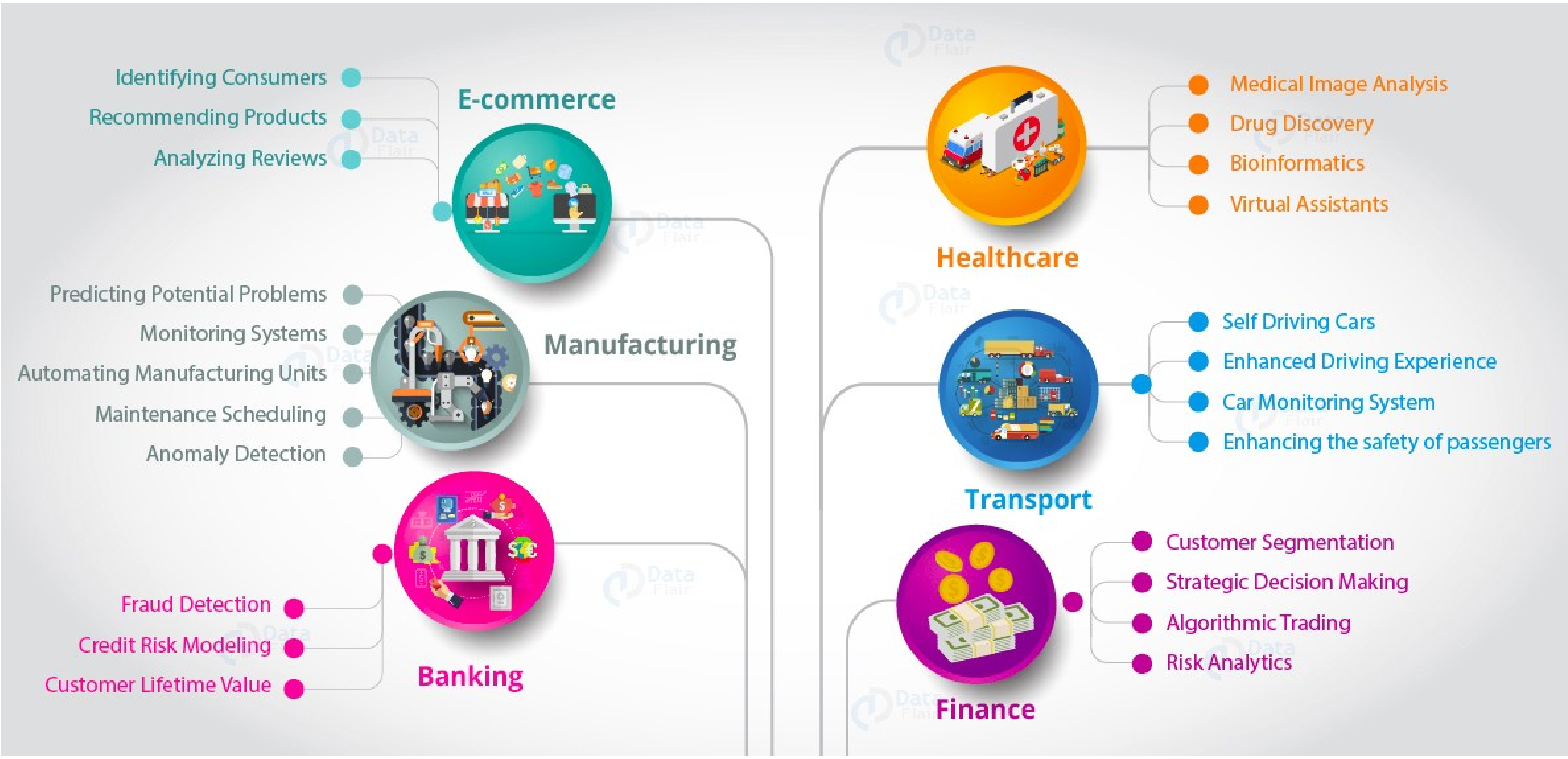
Issues in Machine Learning

- What algorithms exist for learning general target functions from specific training examples ?
- How does the number of training examples influence accuracy ?
- When and how can prior knowledge held by the learner guide the process of generalizing from examples ?

Issues in Machine Learning

- What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem ?
- What is the best way to reduce the learning task to one or more function approximation problems ?
- How can the learner automatically alter its representation to improve its ability to represent and learn the target function ?

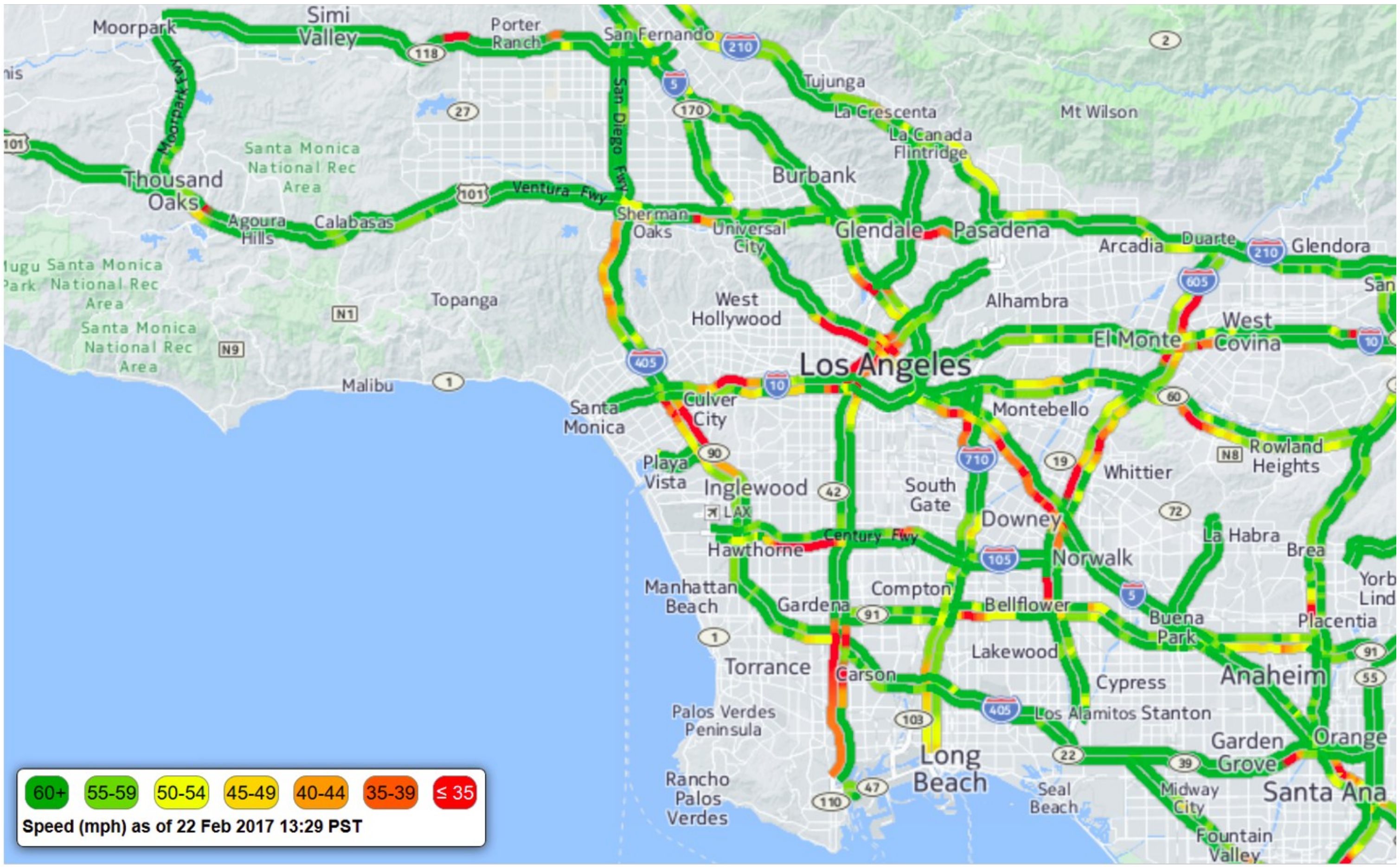
Real Life ML Applications



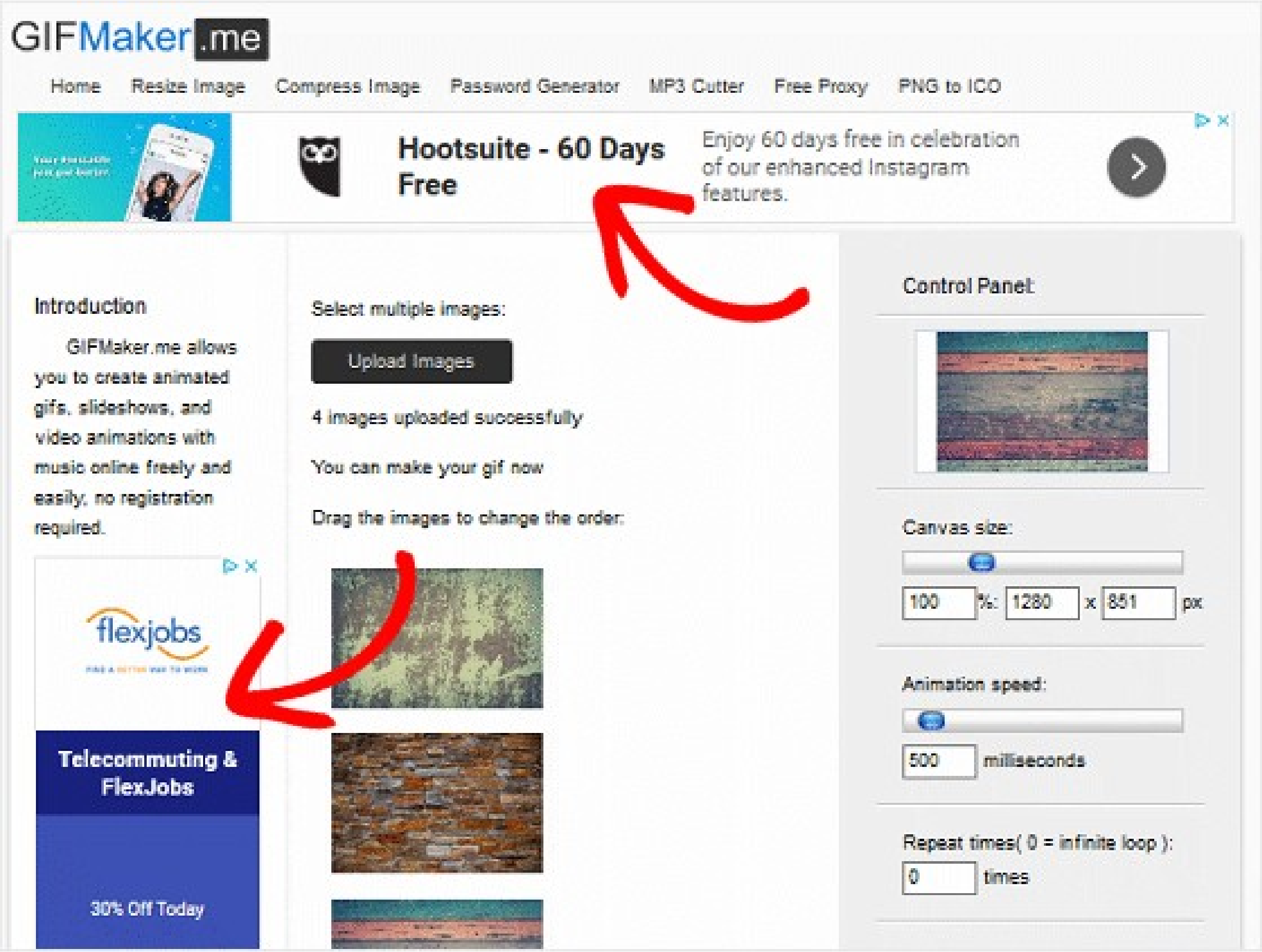
Real Life Examples

- Traffic Prediction
- Targeted Advertising and re-targeting
- Recommendar Systems
- Image Recognition
- Speech Recognition
- Automatic Language Translation
- Self-Driving Cars
- Stock Market Trading
- Fraud and Risk Detection
- Delivery logistics

Traffic Prediction



Targeting Advertisement



Recommender System

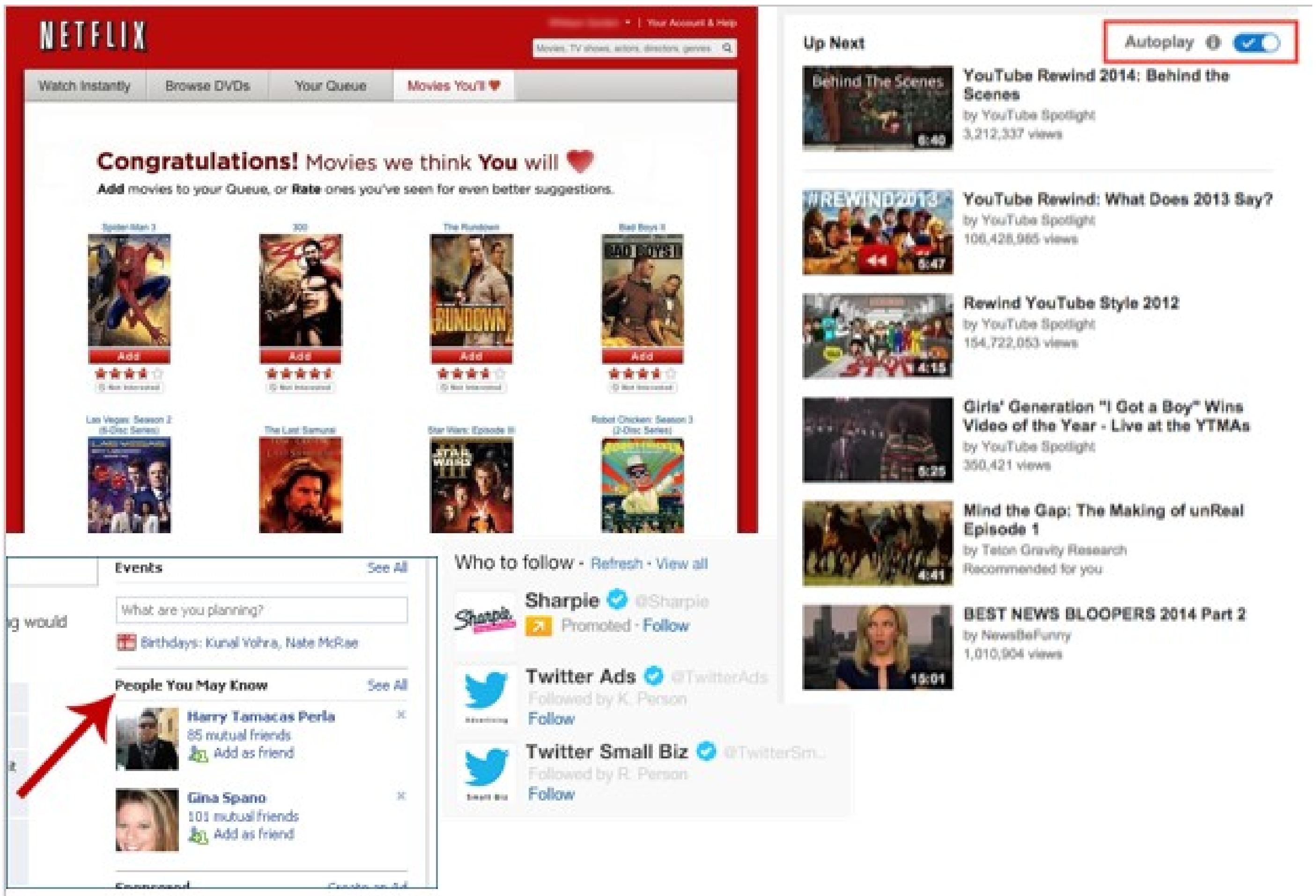
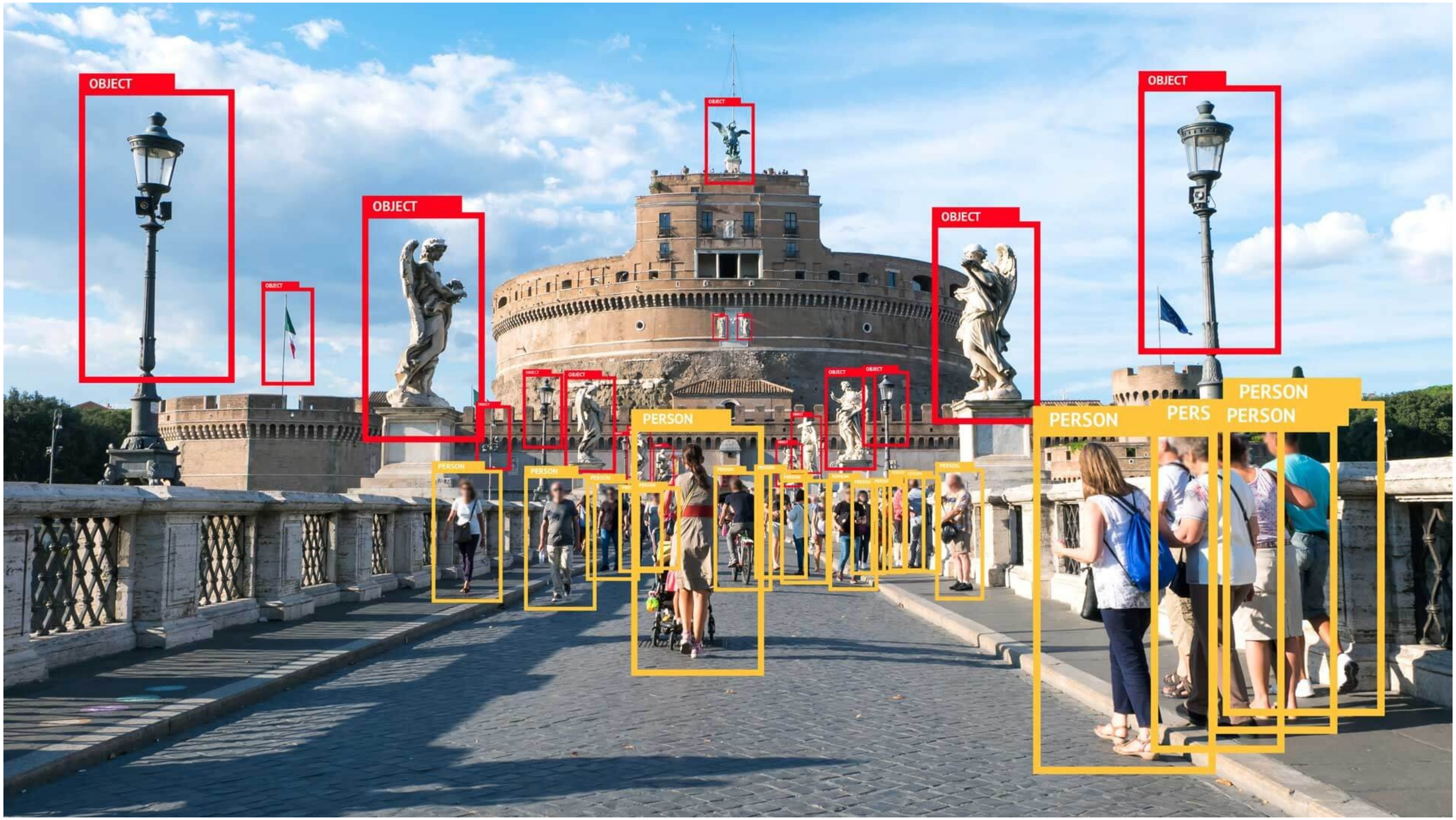
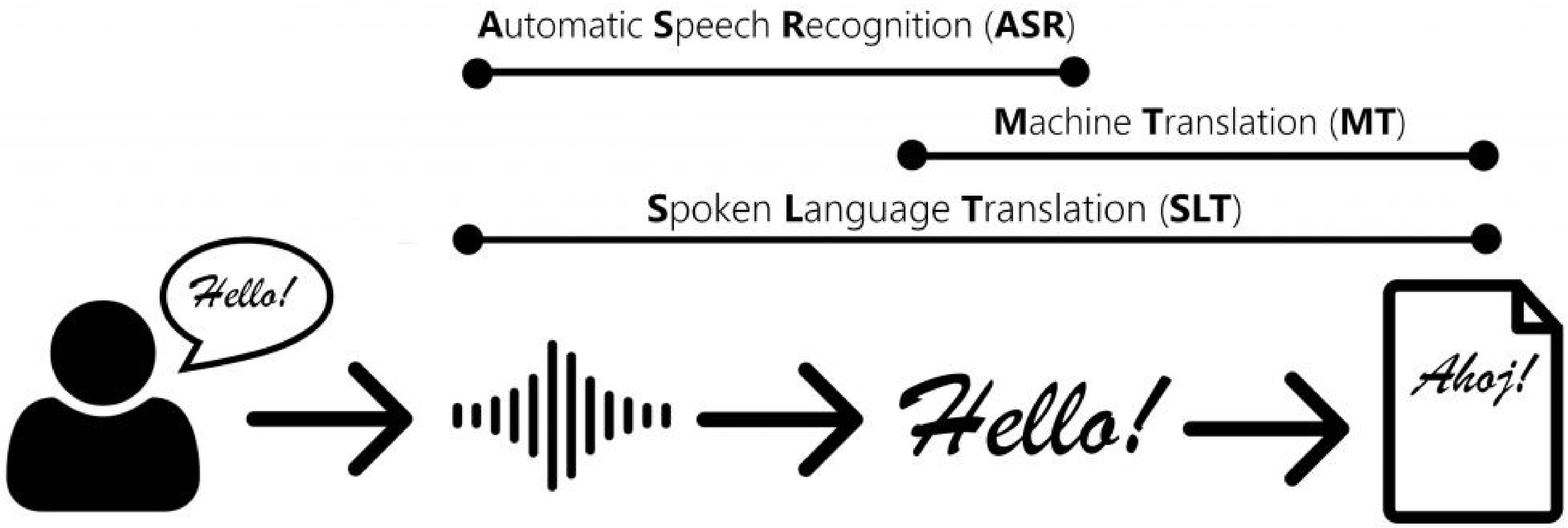


Image Recognition



Speech Recognition

Automatic Language Translation

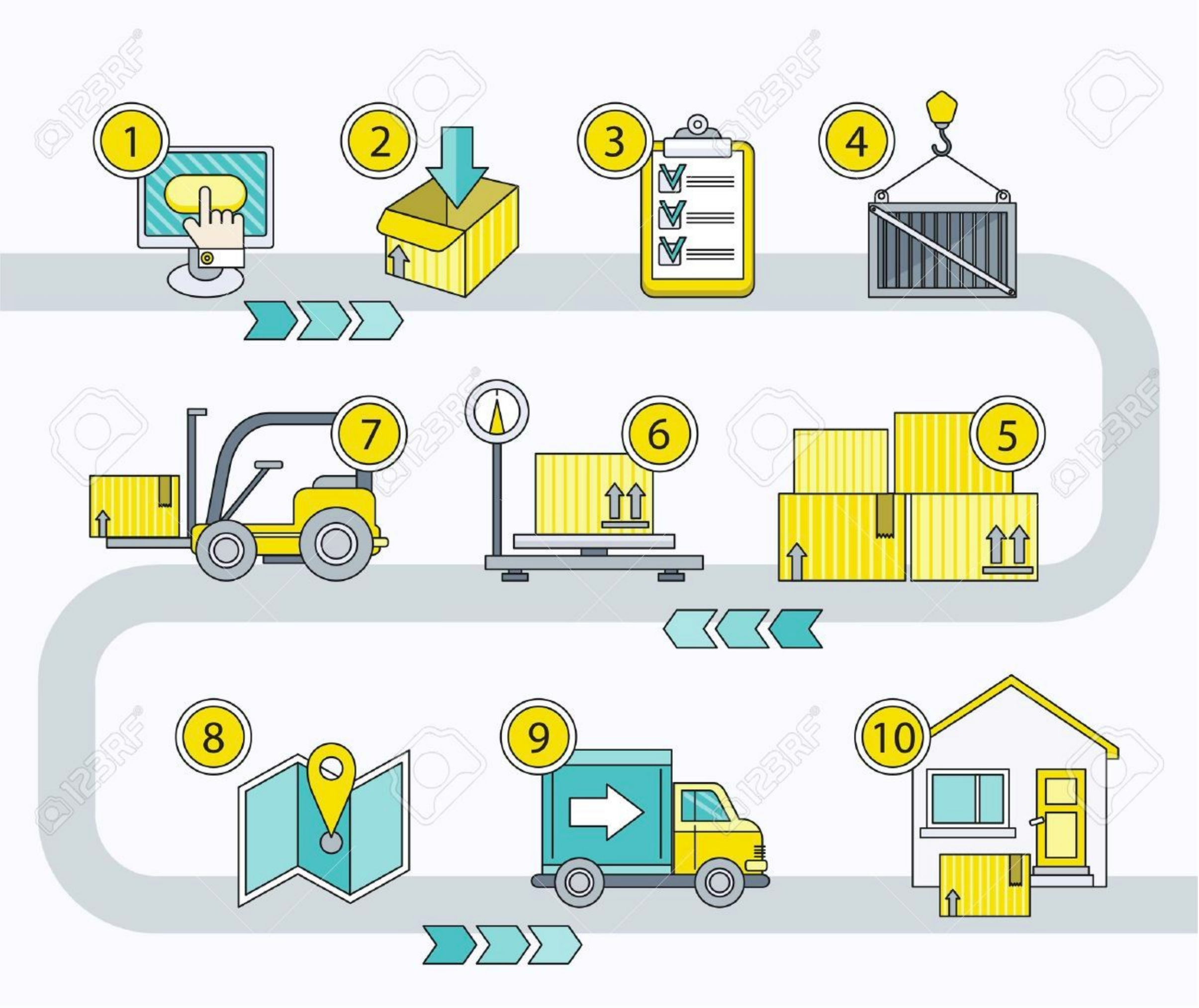


Self-Driving Cars

Stock Market Trading

Fraud Detection

Delivery Logistics



Useful web resources

- www.mitu.co.in
- www.scikit-learn.org
- www.towardsdatascience.com
- www.medium.com
- www.analyticsvidhya.com
- www.kaggle.com
- www.stephacking.com
- www.github.com

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com