

Data Science

An Introduction

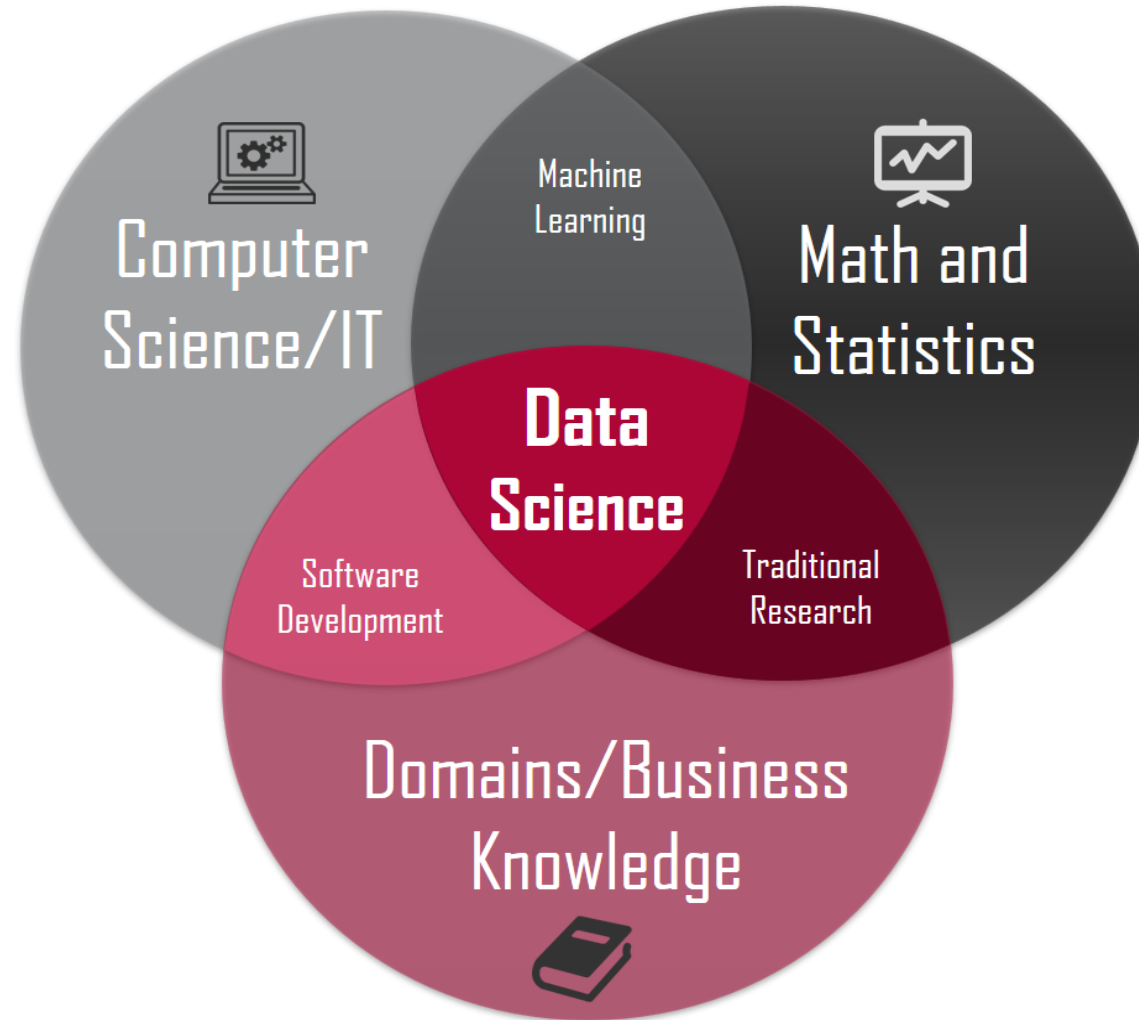
First Let us Understand Terms...

- Data Science
- Analytics
- Artificial Intelligence
- Machine Learning
- Deep Learning

What is Data Science?

- Application of Scientific Methods like Statistical and Machine Learning in order to understand the phenomena to gain control on decision making of it
- It employs techniques from both the fields computer science and statistics
- Data science involves Machine Learning, Segmentation, Visualization and many other things related to data

Data Science Composition



Courtesy: <https://www.fox.temple.edu/institutes-and-centers/data-science/>

What is Analytics?

- Analytics is the discovery, interpretation, and communication of meaningful patterns in data.
- Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance

Types of Analytics



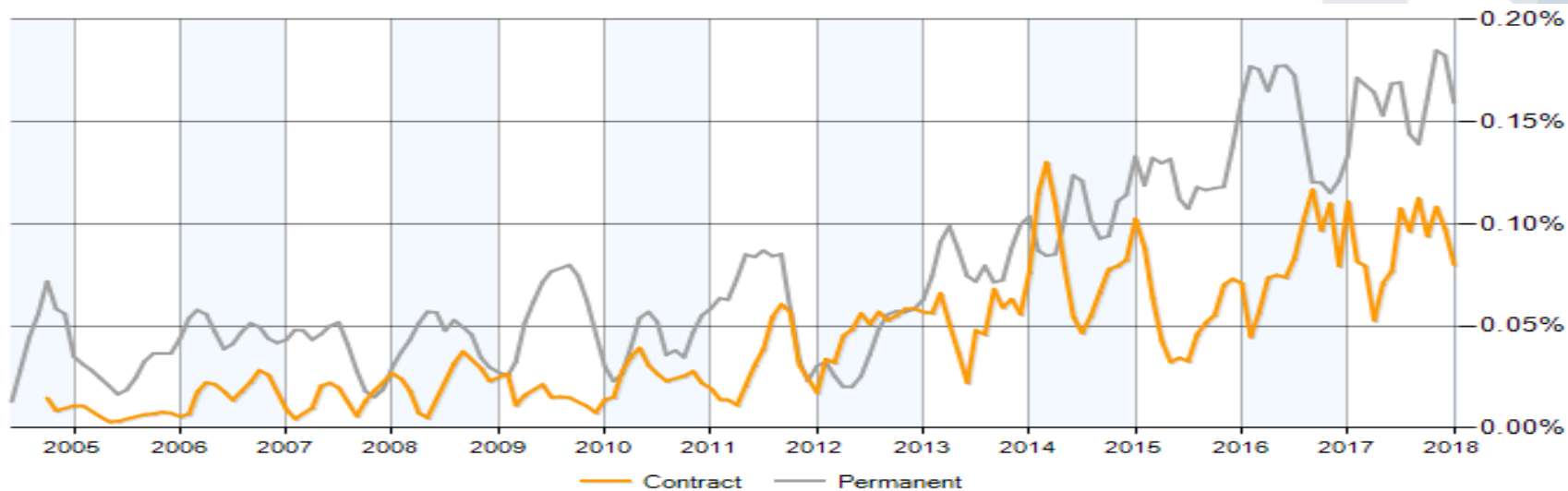
Courtesy: <https://moz.com/blog/when-it-comes-to-analytics-are-you-doing-enough>

Descriptive Analytics



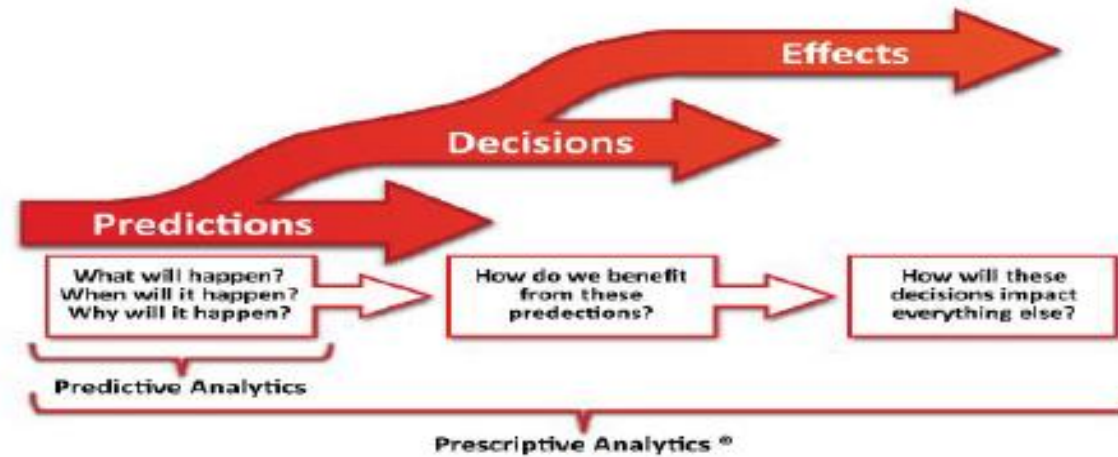
- Gain insight from historical data with reporting, scorecards, clustering etc.
- Can involve data visualization for knowing the basic characteristics of the data
- Descriptive analytics answers the questions what happened and why did it happen.
- Implementations : Business Intelligence, Visualizations
- Software: Informatica, Business Objects, TIBCO Spotfire, Tableau etc.

Predictive Analytics



- Involves statistical and machine learning techniques
- Analyzing the historical patterns in the data and predicting the future patterns
- Predictive analytics answers the question what will happen
- Implementation: Machine Learning, Deep Learning
- Software: R, Python, Libraries like TensorFlow, h2o.ai etc.

Prescriptive Analytics



- Prescriptive analytics goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions and showing the implications of each decision option.
- Implementation: Optimization Techniques like Linear programming Problems, Non-linear programming Problems, Genetic Algorithm etc.

What is AI?

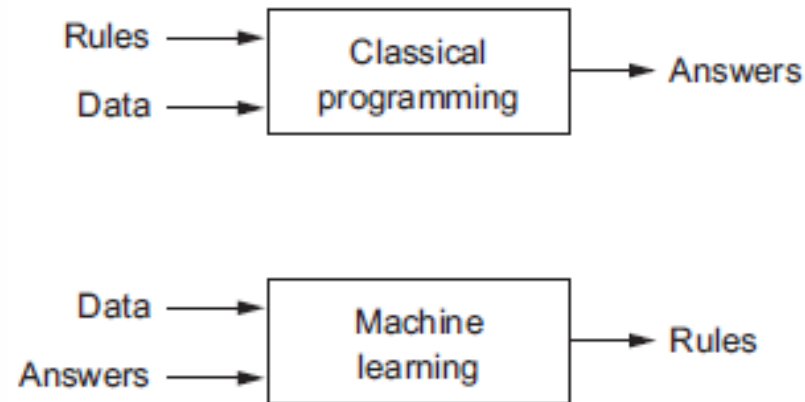
- AI or artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems.
- These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction.

The promise of AI

- Currently, a big hype about AI. Not all realistic
- What it can do perfectly?
 - Answer questions
 - Watch over your health
 - Deliver Groceries at our door
 - Break-through in genomics
- What currently AI cannot do perfectly?
 - Human-level general intelligence
- But AI has achieved a never before height of expectation

What is Machine Learning?

- Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.
- In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed".
- It can be called as “the effort to automate intellectual tasks normally performed by humans”.



Where is Machine Learning Used?

- Medicine: Medical researchers might use it to predict the likelihood of a cancer relapse.
- Intelligence: Intelligence agencies might use it to determine which of a huge quantity of intercepted communications are of interest.
- From a large list of prospective customers, which are most likely to respond?
- To find which customers are most likely to commit fraud?

Role of Machine Learning (ML)

- Machine Learning is used as an aid to achieve AI
- ML Algorithms are driven by mathematical concepts
- ML Algorithms analyse the patterns in the captured data and can be used to build a predictive model on the existing phenomena in business
- Broadly, there are three types of ML Algorithms
 - Supervised Learning Algorithms
 - Unsupervised Learning Algorithms
 - Re-inforcement Learning Algorithms

Supervised Learning

- Supervised learning algorithms are those used in classification and prediction.
- We must have data available in which the value of the outcome of interest (e.g., purchase or no purchase) is known.
- The objective is to predict the values of the outcome of interest

Models for Supervised Learning

- We identify strong links between variables of a data table (columns).
- Such a link may translate into an expression between one variable y (the so-called "dependent" or "response" variable) and a group of other variables $\{x_i\}$ (the so-called "independent variables" or "predictors") :

$$y = f(x_1, x_2, \dots, x_p) + \text{Small random noise}$$

Types in Supervised Learning

- When the response variable is numerical, predictive modeling is called **Regression**.
- When the response variable is categorical (nominal / ordinal), predictive modeling is called **Classification**.

Examples

- **Regression Case:** Sales are influenced by the variables like advertisement expenses, manpower deployed for sales, cost of products, number of dealers etc. Hence we see here
$$\text{Sales} = \text{function}(\text{Adv. Exp} , \text{Manpower} , \text{Cost} , \text{Dealers} , \dots)$$
- **Classification Case:** The customer may purchase a particular product based on some conditions like his need, his age, his income, his place of residence etc. Hence we see here
$$\text{Prob}(\text{Customer Purchases}) = \text{function}(\text{Age}, \text{Income}, \text{Residence}, \dots)$$

Short Quiz: Identify the type

1. An e-commerce company using labeled customer data to predict whether or not a customer will purchase a particular item.
2. A healthcare company using data about cancer tumors (such as their geometric measurements) to predict whether a new tumor is benign or malignant.
3. A factory wanting to predict the time before a break-down of its production machines.
4. A restaurant using review data to ascribe positive or negative sentiment to a given review.
5. A bike share company using time and weather data to predict the number of bikes being rented at any given hour.

Short Quiz: Answers

1. An e-commerce company using labeled customer data to predict whether or not a customer will purchase a particular item. --- **Classification**
2. A healthcare company using data about cancer tumors (such as their geometric measurements) to predict whether a new tumor is benign or malignant. --- **Classification**
3. A factory wanting to predict the time before a break-down of its production machines. --- **Regression**
4. A restaurant using review data to ascribe positive or negative sentiment to a given review. --- **Classification**
5. A bike share company using time and weather data to predict the number of bikes being rented at any given hour. --- **Regression**

Algorithms of Supervised Learning

- Naïve Bayes
- K-NN
- Decision Trees
- Regression Models
- Neural Nets
- Support Vector Machines

Unsupervised Learning

- Unsupervised learning algorithms are those used where there is no outcome variable to predict or classify.
- These type of methods are used many times for *exploratory data analysis*
- Association rules, data reduction methods and clustering techniques are all unsupervised learning methods.

Examples

- Customer Segmentation like RFM (Recency, Frequency, Monetary)
- Market Basket Analysis
- Product Grouping

Algorithms of Unsupervised Learning

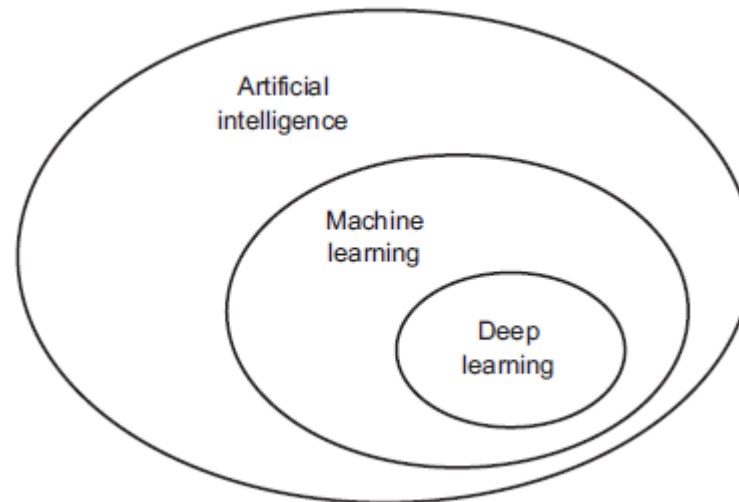
- Clustering Techniques
 - Hierarchical
 - K-means
 - K-medoids
- Principal Component Analysis
- Association Rules

Re-inforcement Learning

- In this type, there is an agent which/who receives information from the environment and learns to choose actions based on rewards or punishment received
- Examples include:
 - Self-driving cars
 - Robotics
- Algorithms:
 - Upper Confidence Bound
 - Thomson Sampling

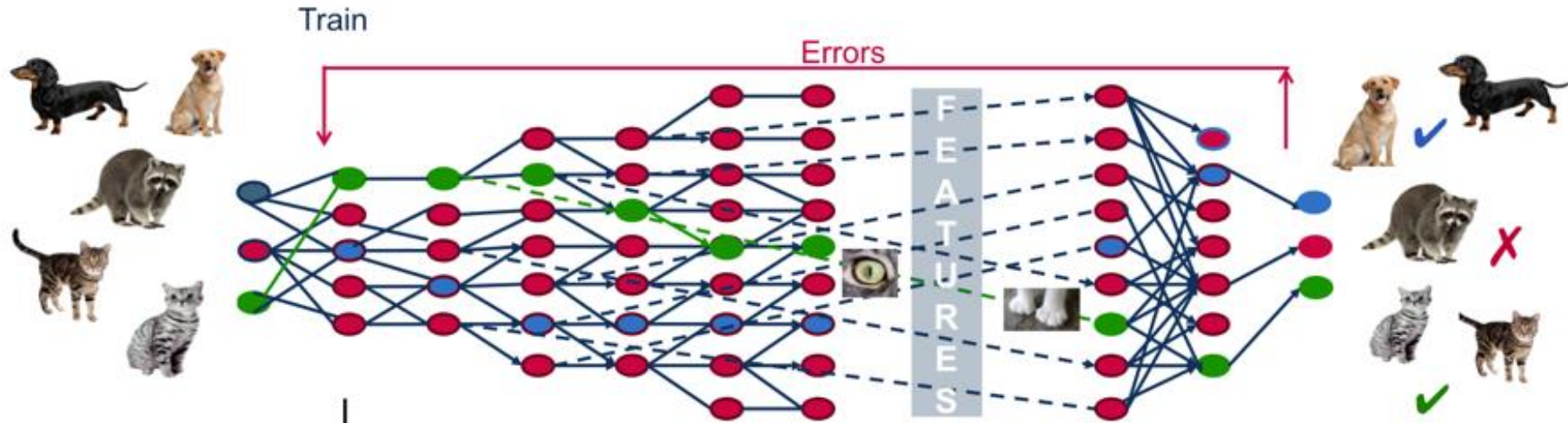
What is Deep Learning?

- Deep learning is a specific subfield of machine
- The *deep* in deep learning isn't a reference to any kind of deeper understanding achieved by the approach; rather, it stands for this idea of successive layers of representations.
- The concept made use of in deep learning is of *neural network* algorithm

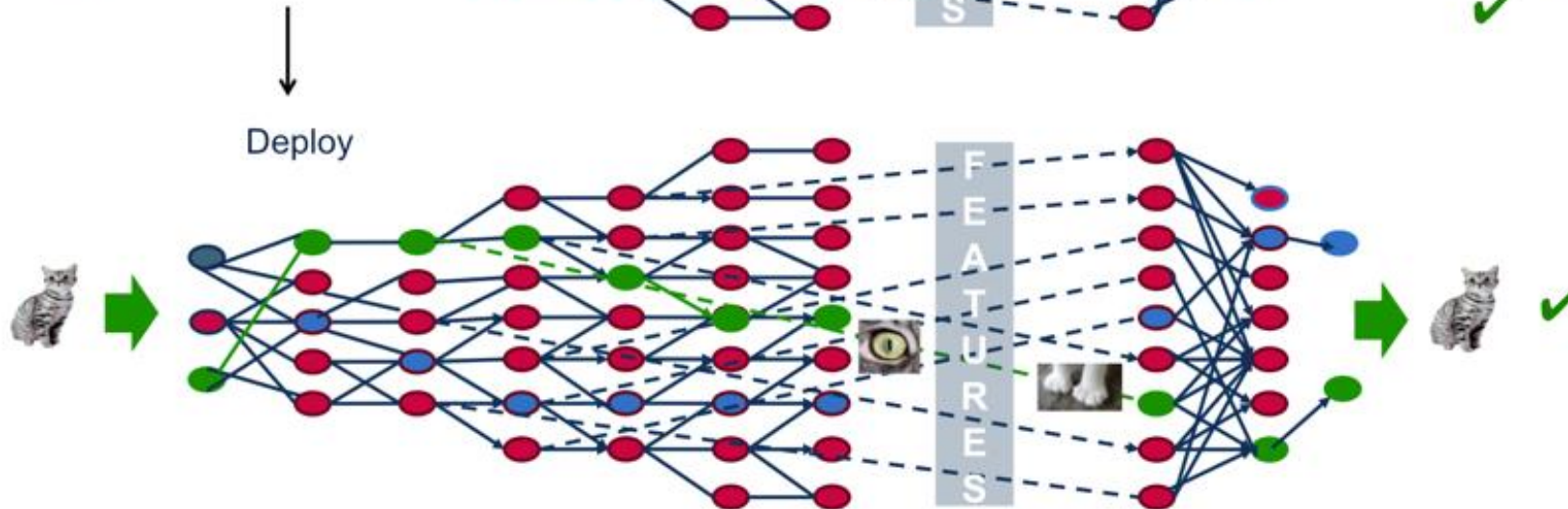


How does Deep Learning Work?

Training an algorithm to identify a **cat**

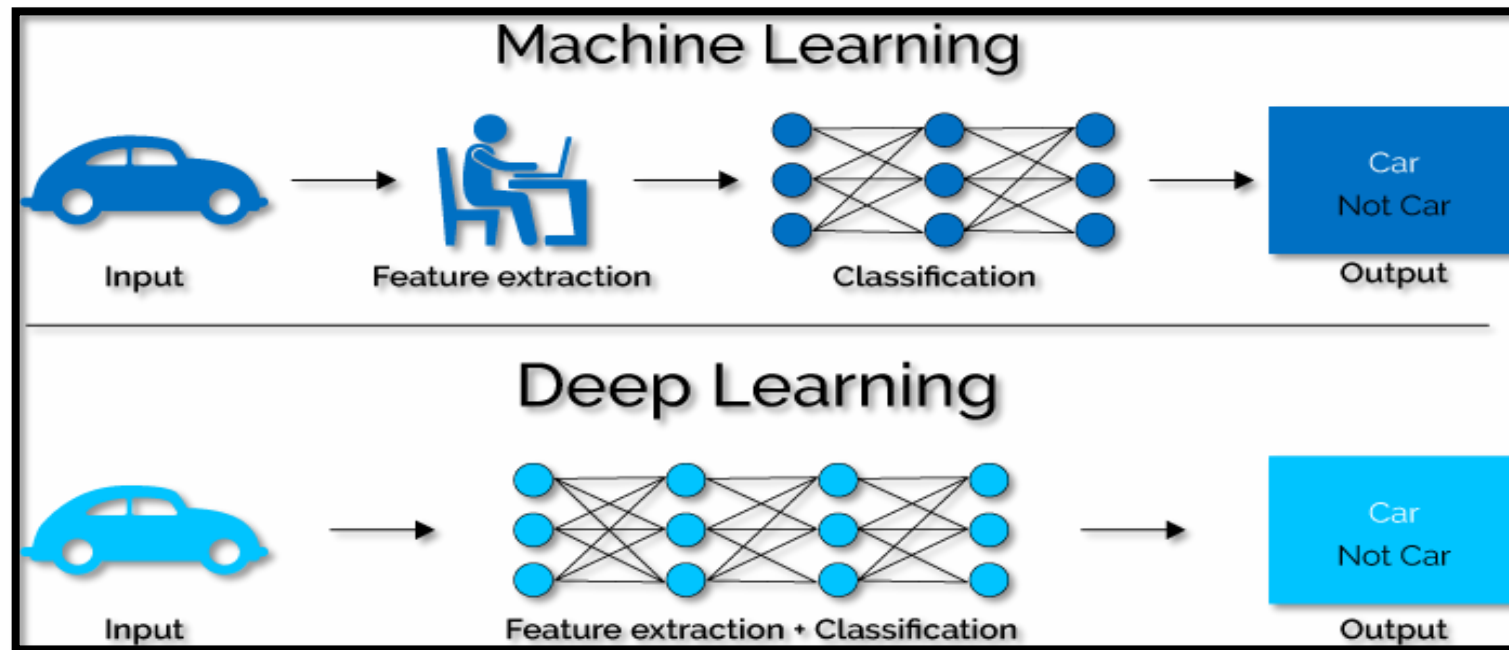


Identifying a **cat**



What Deep Learning provides which ML doesn't?

- Surpasses the crucial step in ML i.e. feature extraction
- Effective on complex problems like image and voice recognition
- Allows a model to learn all layers of representation jointly



Courtesy: <https://www.xenonstack.com/blog/data-science/log-analytics-with-deep-learning-and-machine-learning>

Questions ?