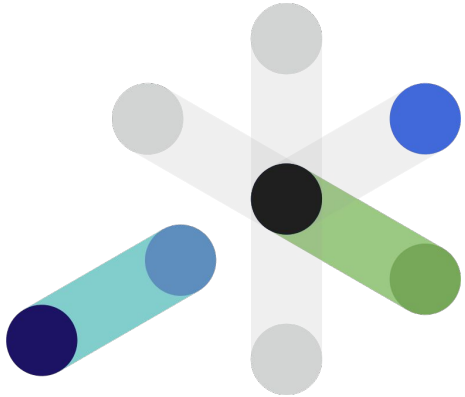
srijan:

Machine Learning 101 and path to career in data science



By: Mayank Kumar

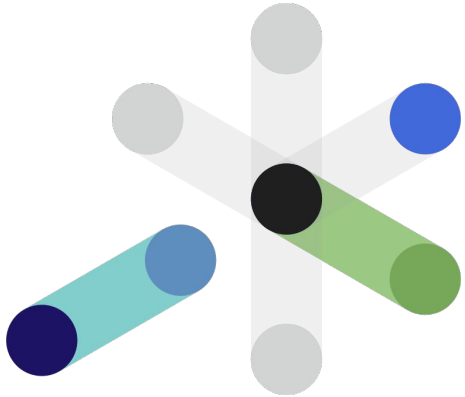
About Me

Mayank Kumar

Data Scientist - II

Experience Across
Machine Learning, Deep Learning,
MLOps, Cloud, Algorithms, Optimization

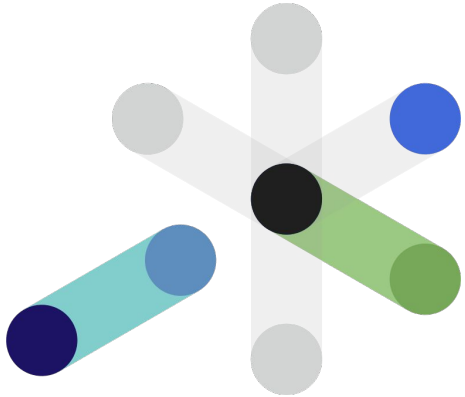
srijan:



LinkedIn : https://www.linkedin.com/in/mayank-k-jha/

srijan:

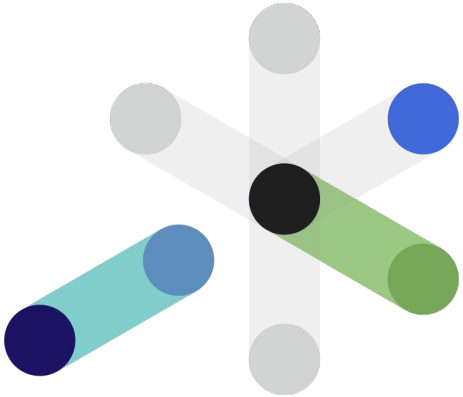
Part 1: Machine Learning 101



Things to cover

srijan:

1. Introduction to Machine Learning
2. Basic Machine Learning Terminologies
3. Machine Learning Approaches
4. Process of Bootstrapping a Machine Learning Project



Introduction to Machine Learning

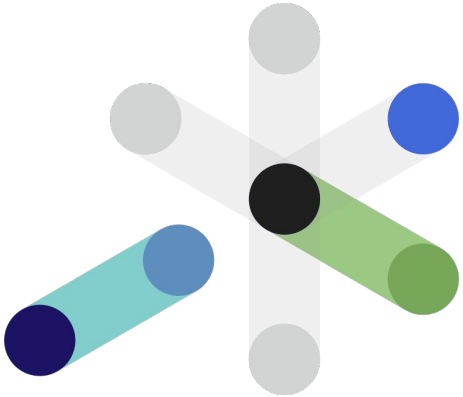
srijan:

What is Machine Learning ?

- “Learning is any process by which a system improves performance from experience.” - **Herbert Simon**

- **Definition by Tom Mitchell (1998)**: Machine Learning is the study of algorithms that
  - Improve their performance P
  - At some task T
  - With experience E.

  A well-defined learning task is given by <P, T, E>

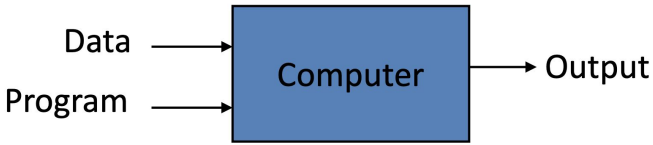


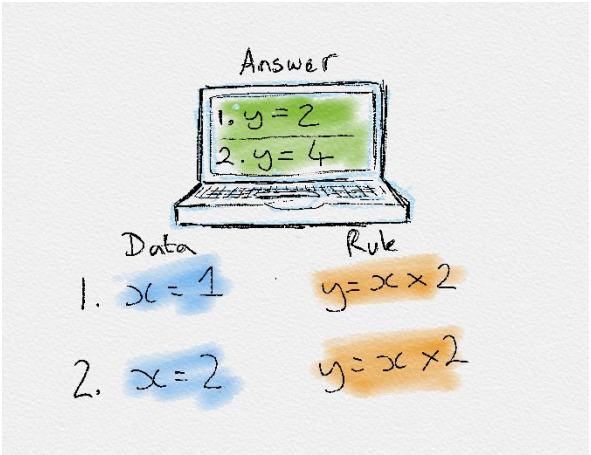
Introduction to Machine Learning

srijan:

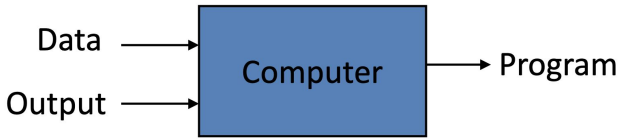
What is Machine Learning ?

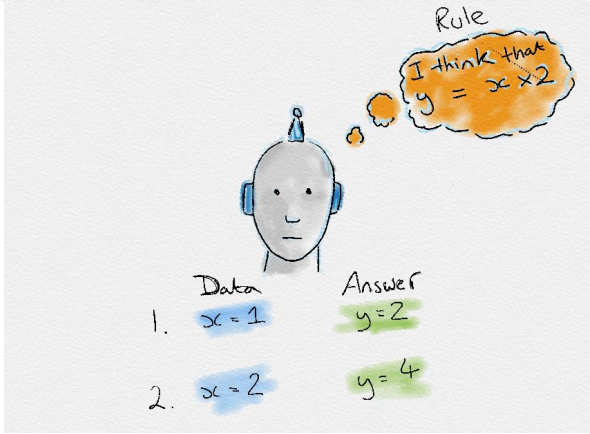
Traditional Programming

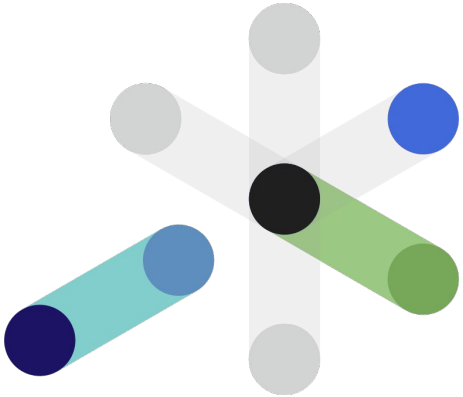




Machine Learning







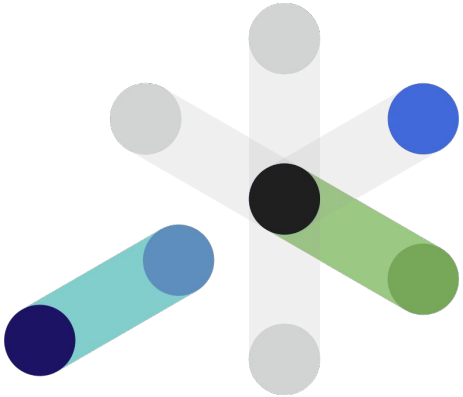
ref: https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0

Introduction to Machine Learning

srijan:

What is Machine Learning ?

- Some examples where Machine Learning can be used :
  - Humans can’t explain their expertise (speech recognition)



- Some examples where Learning isn’t always useful:
  - There is no need to “learn with experience” when calculating a number is prime number or not

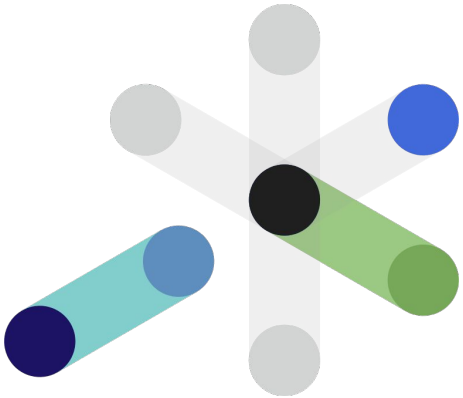
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Introduction to Machine Learning

srijan:

What is Machine Learning ?

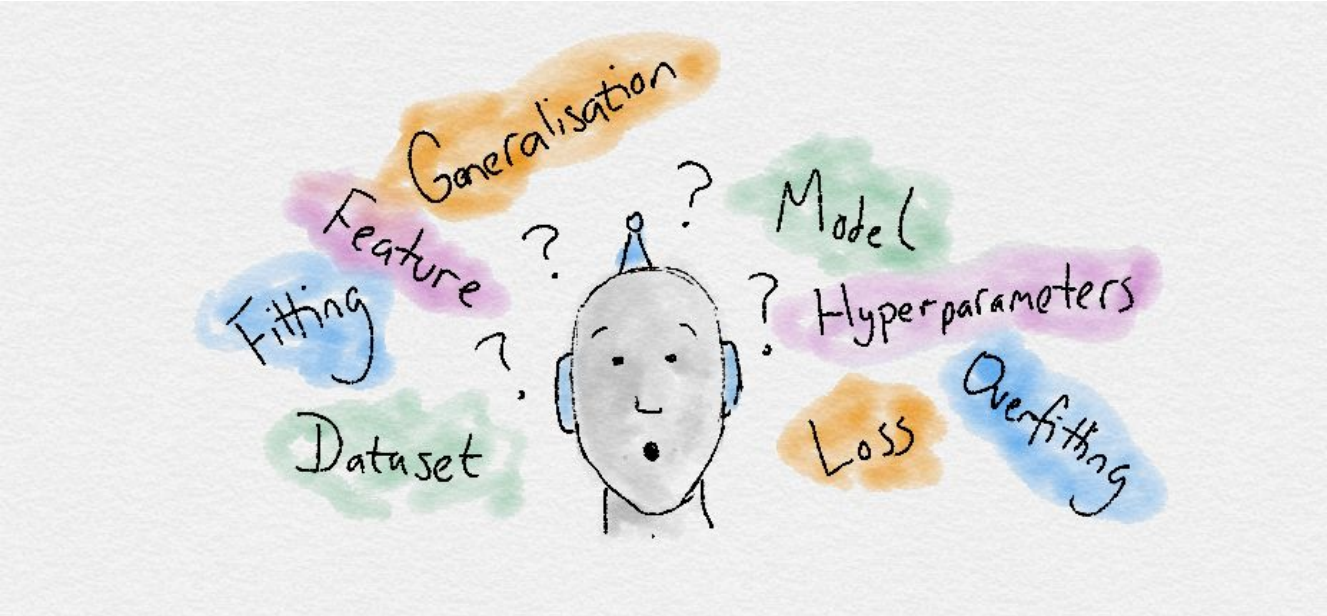
- Some more examples of tasks that are best solved by using a learning algorithm:
  - Recognizing patterns :
    - Facial identities or facial expressions
    - Handwritten or spoken words Recognition
    - Medical images (ex: detecting brain Tumors)
  - Recognizing anomalies :
    - Unusual credit card transactions
    - Unusual patterns of sensor readings in a nuclear power plant
  - Prediction :
    - Future stock prices or currency exchange rates

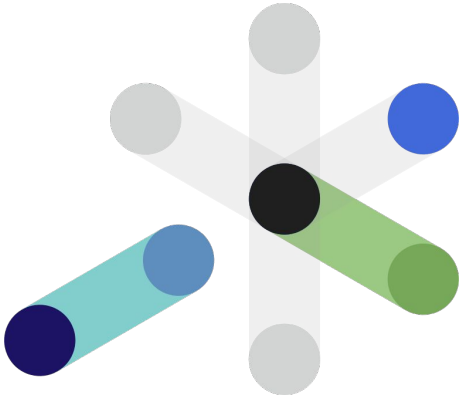


ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Basic Machine Learning Terminologies

srijan:





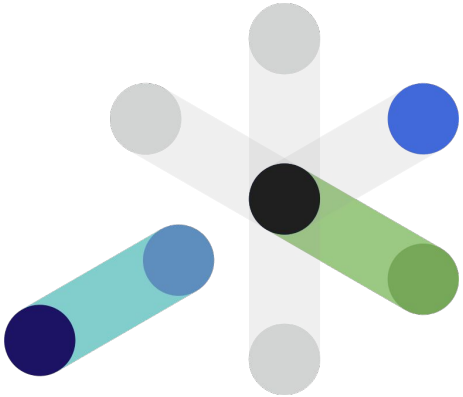
ref: https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0

Basic Machine Learning Terminologies

srijan:

- Most commonly used simple terminologies are:
  - Dataset: A set of data examples, that contain features important to solving the problem.

	name	age	score
0	Jhon	28	3.75
1	David	55	9.50
2	Adam	19	5.70
3	Sara	46	7.60

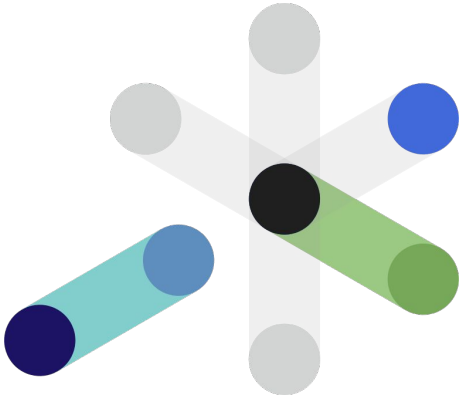


Basic Machine Learning Terminologies

srijan:

- Most commonly used simple terminologies are:
    - Features: An individual measurable property or characteristic of a phenomenon being observed.

	name	age	score
0	Jhon	28	3.75
1	David	55	9.50
2	Adam	19	5.70
3	Sara	46	7.60



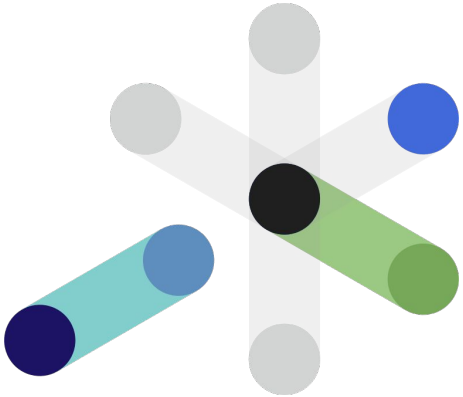
For example, in above dataset, name, age can be called as features for determining scores.

Basic Machine Learning Terminologies

srijan:

- Most commonly used simple terminologies are:
  - Target: A known value for a given phenomenon being observed.

	name	age	score
0	Jhon	28	3.75
1	David	55	9.50
2	Adam	19	5.70
3	Sara	46	7.60

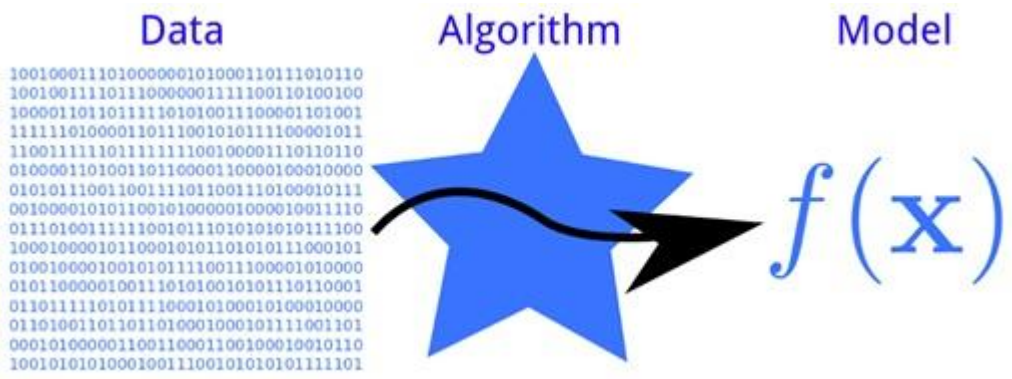


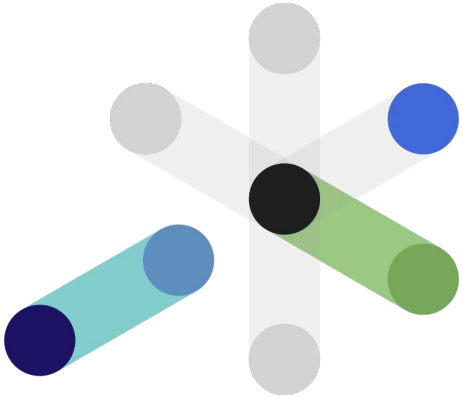
For example, in above dataset, score can be called as target

Basic Machine Learning Terminologies

srijan:

- Most commonly used simple terminologies are:
  - Model: A mathematical representation of a real-world process



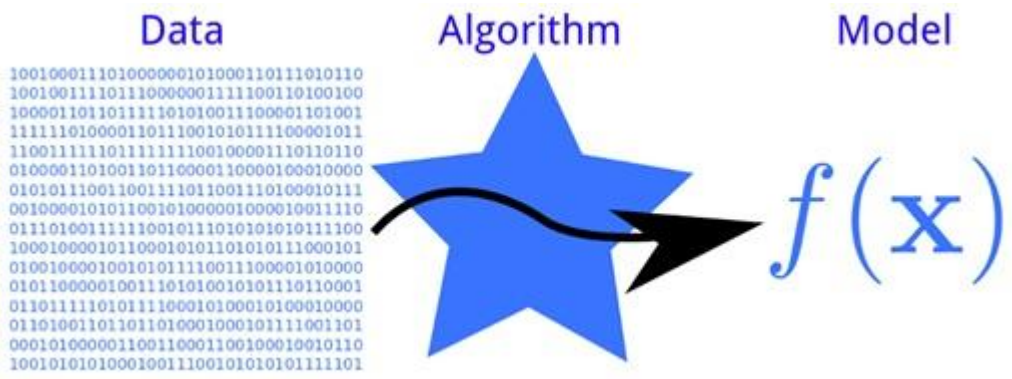


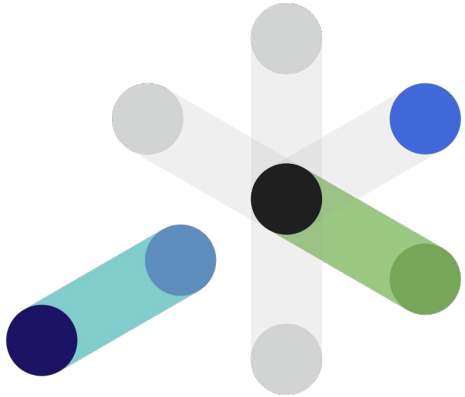
ref: https://www.datasciencecentral.com/profiles/blogs/how-to-choose-a-machine-learning-model-some-guidelines

Basic Machine Learning Terminologies

srijan:

- Most commonly used simple terminologies are:
  - Training: Process of learning a mathematical function say f(x) from the given dataset.



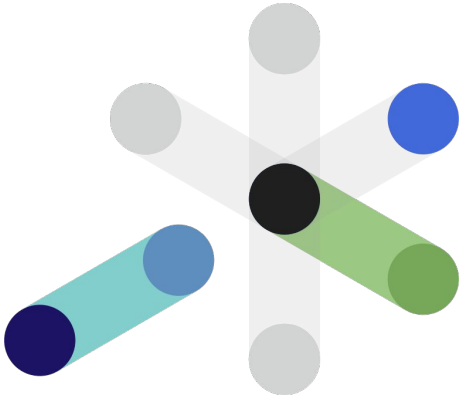


ref: https://www.datasciencecentral.com/profiles/blogs/how-to-choose-a-machine-learning-model-some-guidelines

Machine Learning Approaches

srijan:

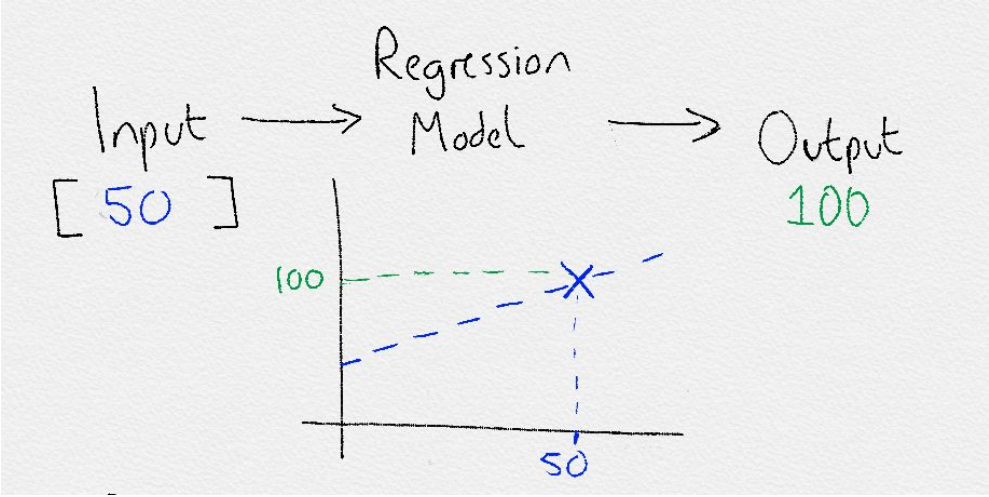
- Following are the various types of Machine Learning problems:
  - Supervised learning – Given:training data + desired outputs (labels)
    - Classification Problem
    - Regression Problem
  - Unsupervised learning – Given: training data (without desired outputs)
  - Reinforcement learning – Rewards from sequence of actions

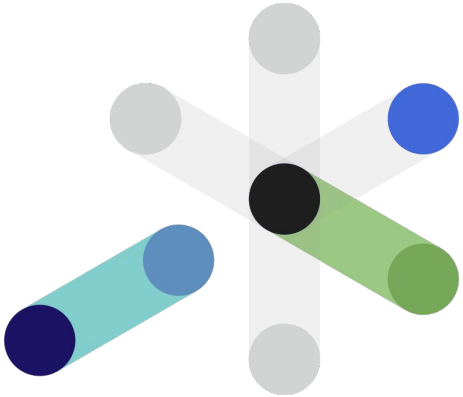


Machine Learning Approaches

srijan:

- Supervised learning – Regression Problem
  - Regression models are used to predict a continuous value.
  - Example:
    - Predicting prices of a house given the features of house like size, area etc.
    - Predicting the score of students in an exam



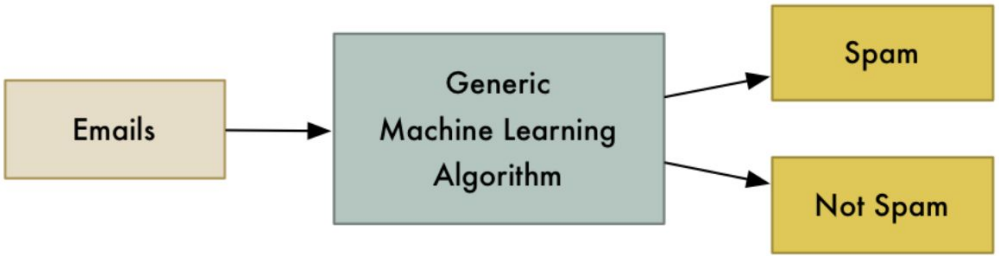


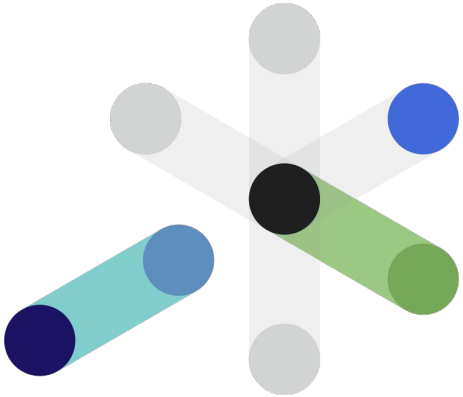
ref: https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0

Machine Learning Approaches

srijan:

- Supervised learning – Classification Problem
  - Classification models are used to predict a categorical values or classes for any event.
  - Example:
    - Predicting an email is spam or legitimate.



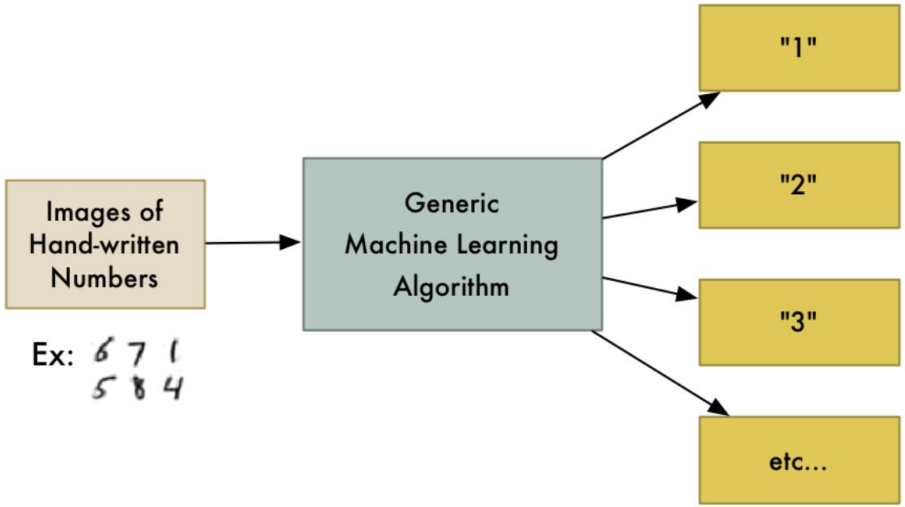


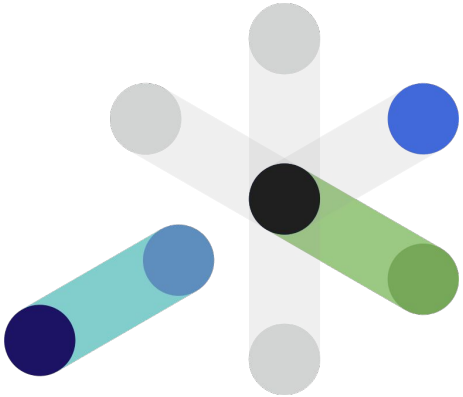
ref: https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471

Machine Learning Approaches

srijan:

- Supervised learning – Classification Problem
  - Classification models are used to predict a categorical values or classes for any event.
  - Example:
    - Handwritten digit recognition



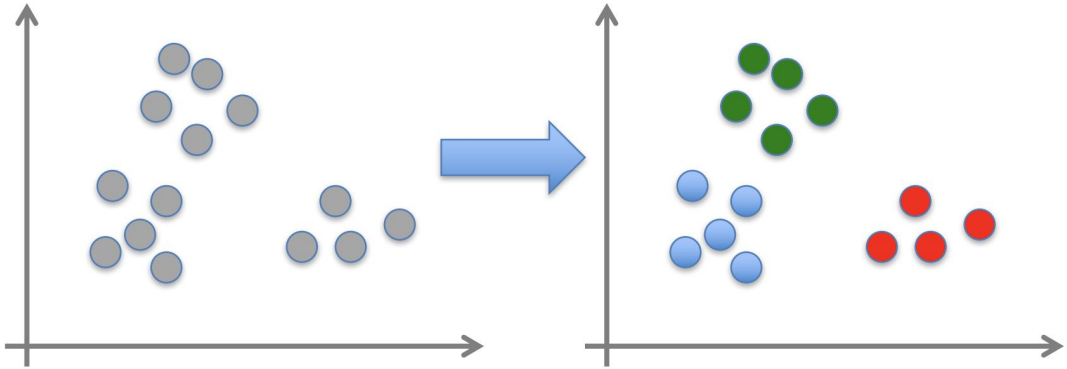


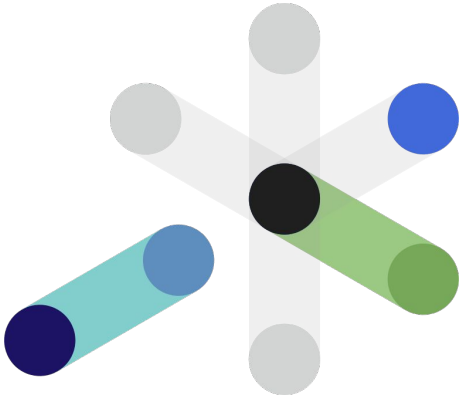
ref: https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471

Machine Learning Approaches

srijan:

- Unsupervised learning
  - No targets/labels are provided in the data
  - Used for finding structures and interesting patterns in data
  - Example:
    - Cluster Analysis:
      - Used to group or categorize the data into some useful patterns
      - Find commonalities in data on the basis of which the grouping is performed.



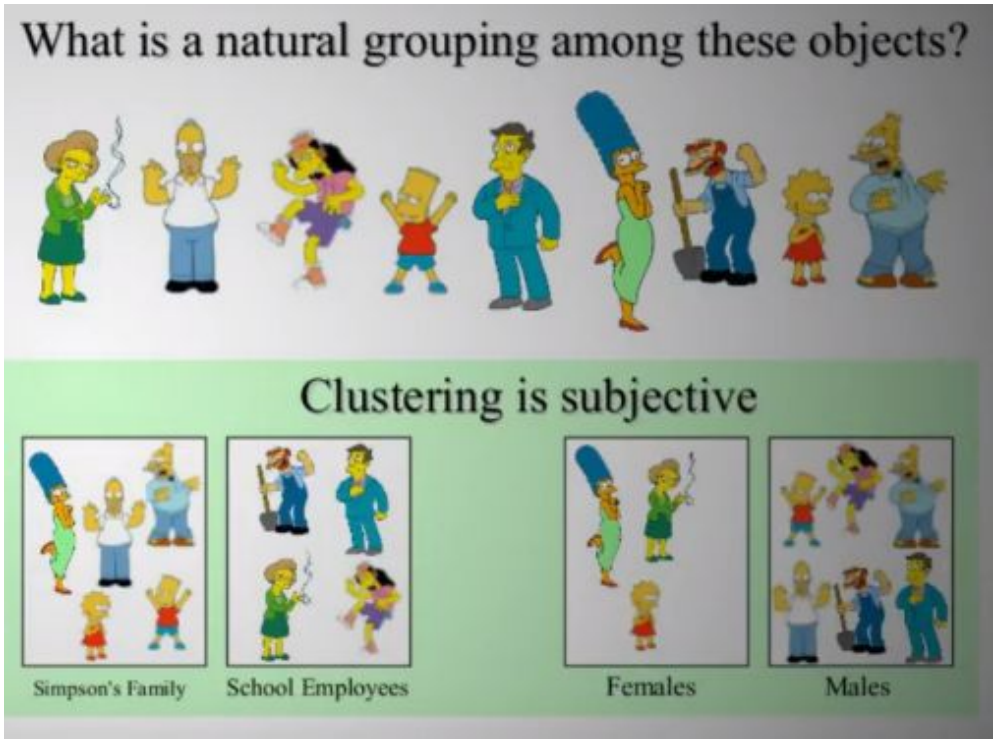


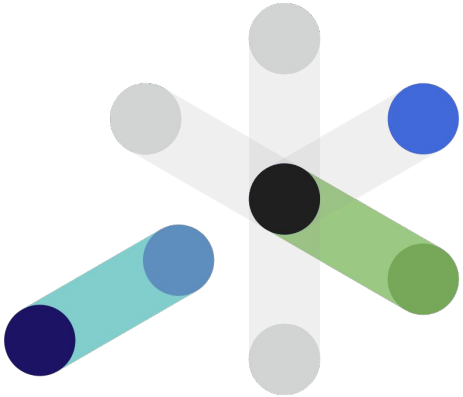
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

srijan:

- Unsupervised learning
  - A simple example





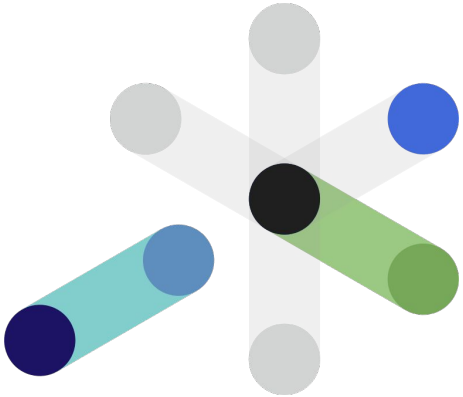
ref: https://medium.com/@randylaosat/a-beginners-guide-to-machine-learning-dfadc19f6caf

Machine Learning Approaches

srijan:

- Unsupervised learning
  - Some real life examples can be

Organize computing clusters



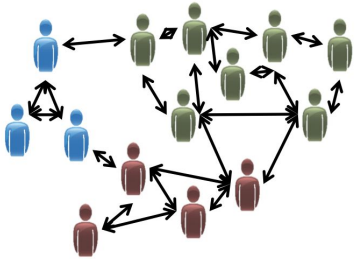
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

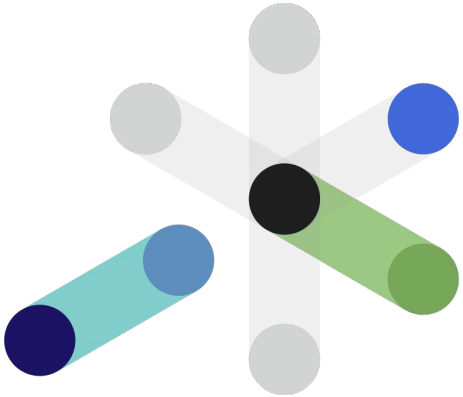
srijan:

- Unsupervised learning
  - Some real life examples can be

Organize computing clusters



Social network analysis



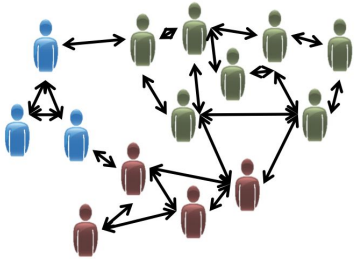
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

srijan:

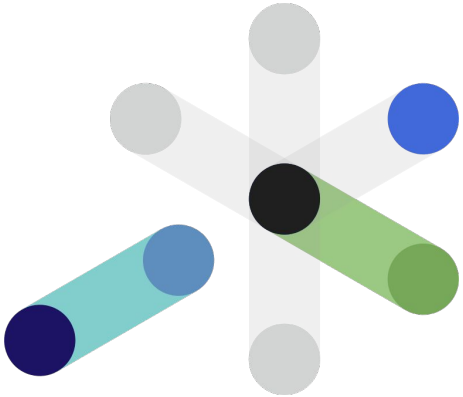
- Unsupervised learning
  - Some real life examples can be

Organize computing clusters



Social network analysis

Market segmentation



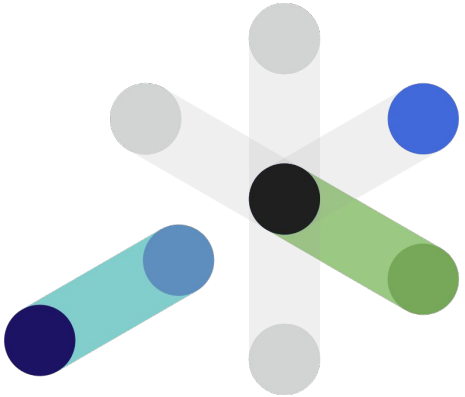
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

srijan:

- Reinforcement learning
  - Given a sequence of states and actions with rewards, output a policy
  - Policy is a mapping from states to actions that tells you what to do in a given state

  - Examples: –
    - Credit assignment problem
    - Game playing
    - Robot in a maze
    - Balance a pole on your hand

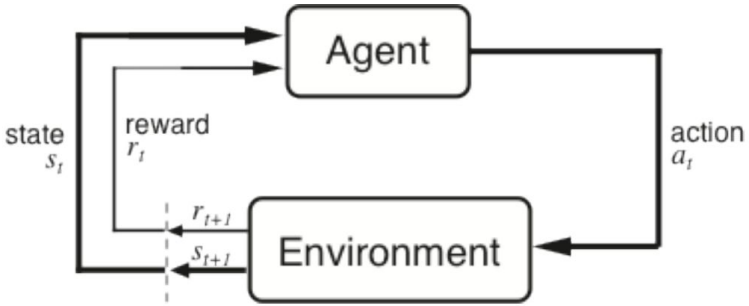


ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

srijan:

- Reinforcement learning



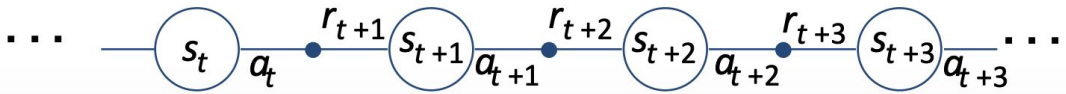
Agent and environment interact at discrete time steps : $t = 0, 1, 2, \mathrm{K}$

Agent observes state at step t: $s_t \in S$

produces action at step t: $a_t \in A(s_t)$

gets resulting reward : $r_{t+1} \in \Re$

and resulting next state : s_{t+1}

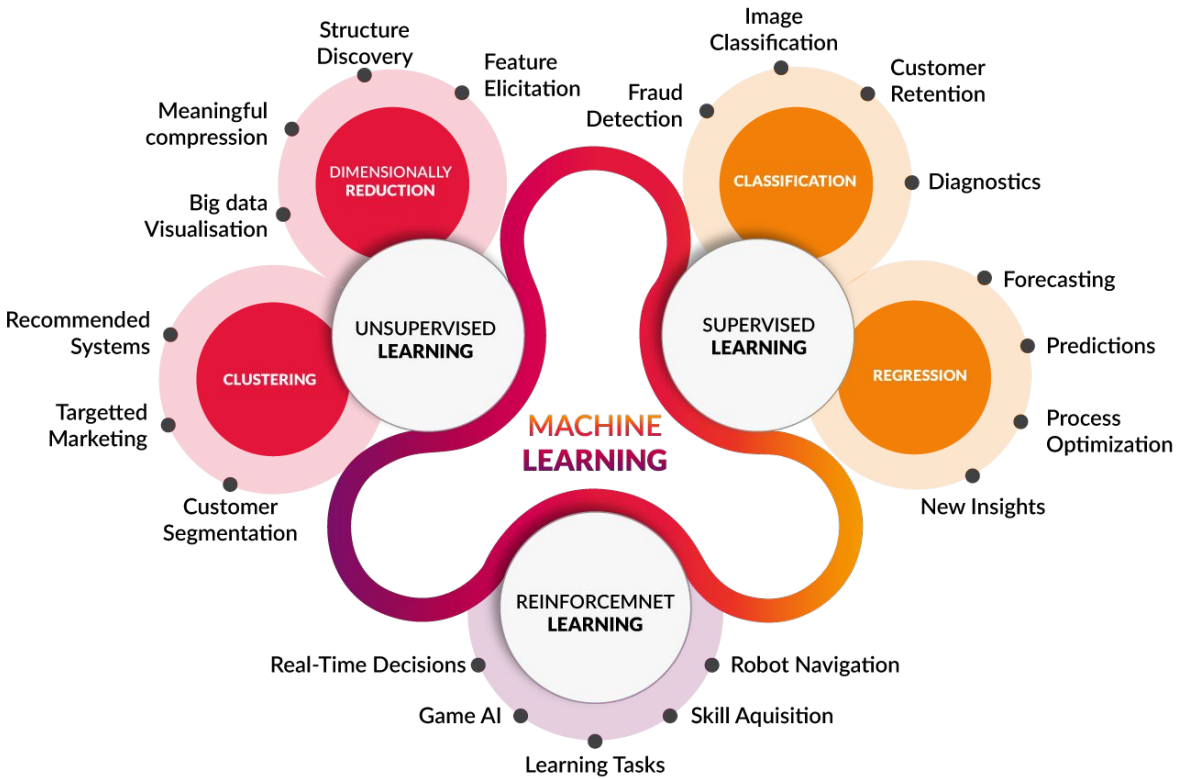


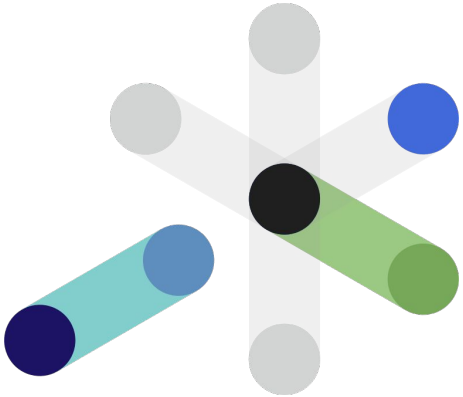
ref: https://www.seas.upenn.edu/~cis 519/fall2017/lectures/01_introduction.pdf

Machine Learning Approaches

srijan:

- Summarizing all approaches with real life applications

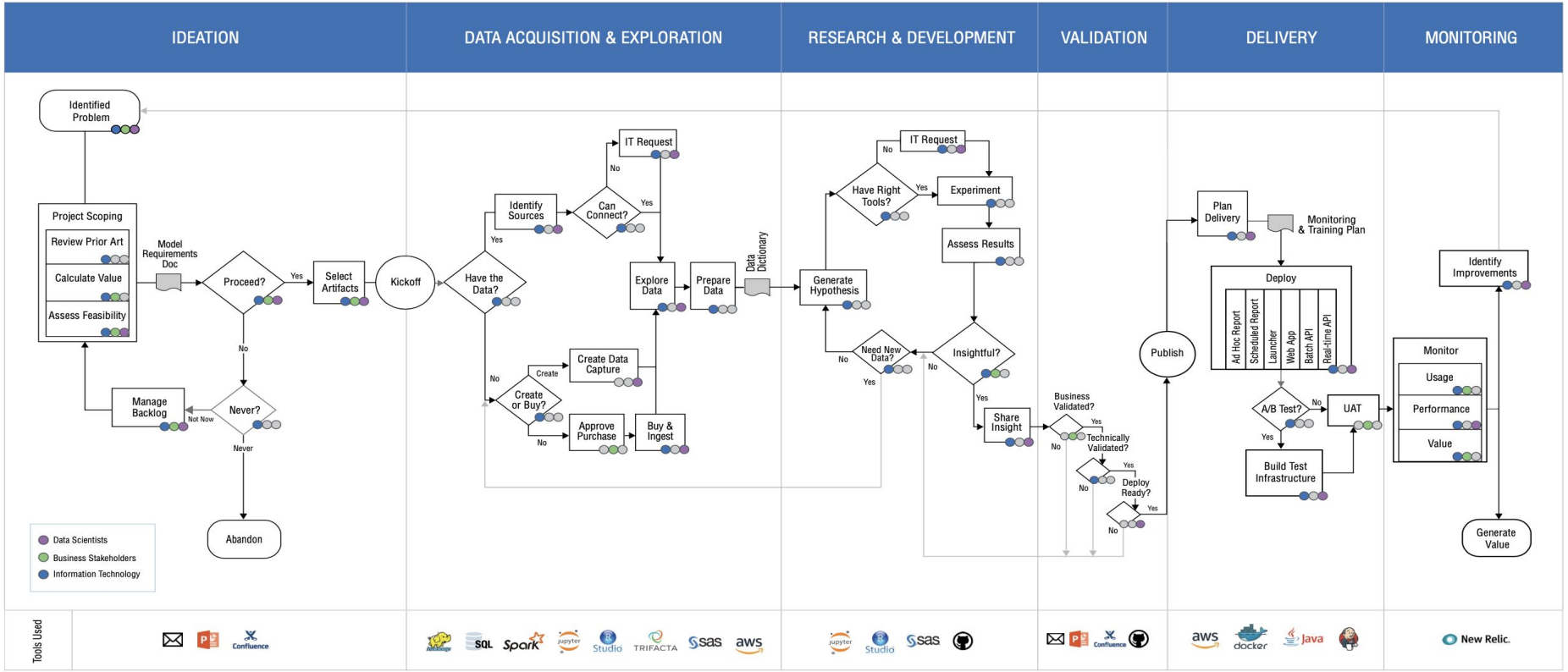




ref: https://www.quora.com/Which-big-problems-does-machine-learning-solve-for-businesses

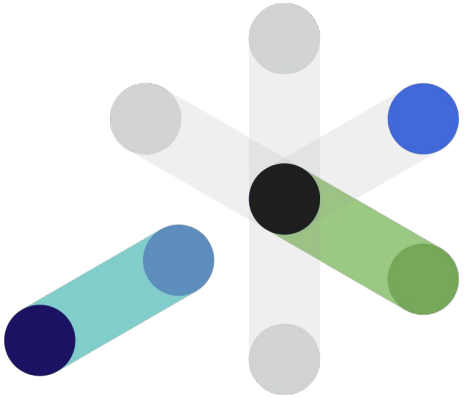
Process of Bootstrapping a Machine Learning Project

srijan:



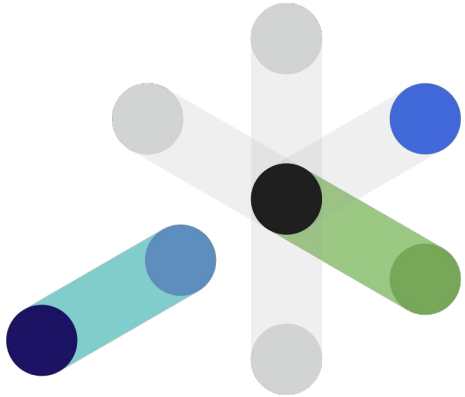
Source: Domino Data Lab

Any Questions ?



srijan:

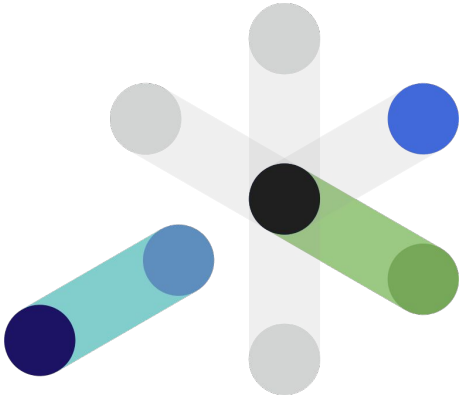
Part 2: Path to career in Data Science



Things to cover

srijan:

1. How to get started with data science
2. Specialization domains
3. How to prepare for data science interviews



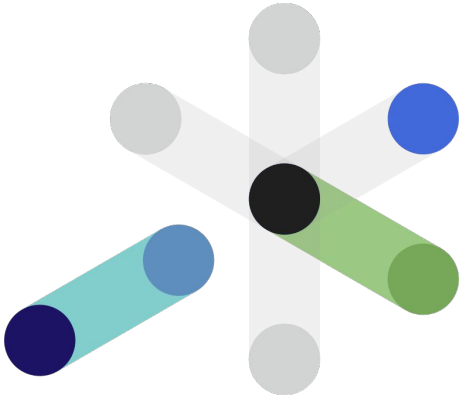
How to get started with data science

srijan:

Step 1:

Be good at coding

- Learn data structure
- Learn algorithm design (focus on why it works?)
- Start with competitive coding
- Bonus (earn some good achievements to prove, you are good at algorithm design and coding)



Benefits:

- It will save tons of your time for writing a production grade system and that will give you more edge over other candidates.

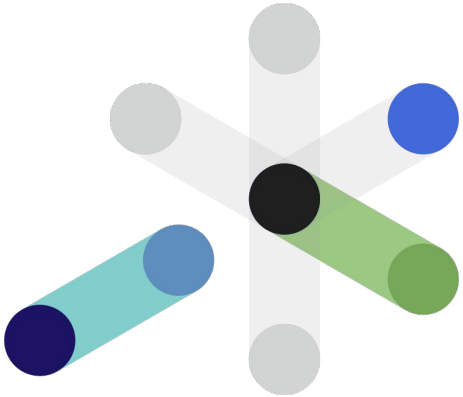
How to get started with data science

srijan:

Step 2:

Be good at mathematics / Academics

- Learn M1, M2, M3, M4 (focus more on why it works?)
- Implement some techniques
- Bonus (Being good at academics will add a lot of advantage)



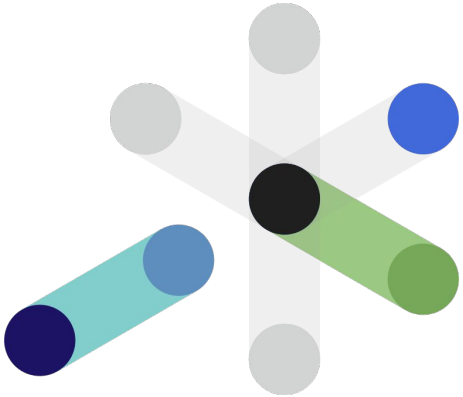
How to get started with data science

srijan:

Step 3:

Start with reading articles about data science and ML

- Make notes of all the basics
- Recommended source: Analytics vidhya, KDNuggets, Medium, Topbots
- Follow DFS approach



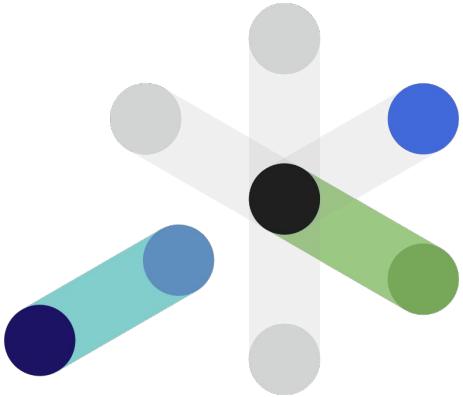
How to get started with data science

srijan:

Step 4:

Start with below courses

- Introduction to ML - Andrew NG
- Probability and Stats - Khan Academy
- Hands on ML in python (Do hands on practise) - Sentdex
- Make notes of all the basics
- Start with sklearn documentation, explore it and learn algorithms
- Follow DFS approach



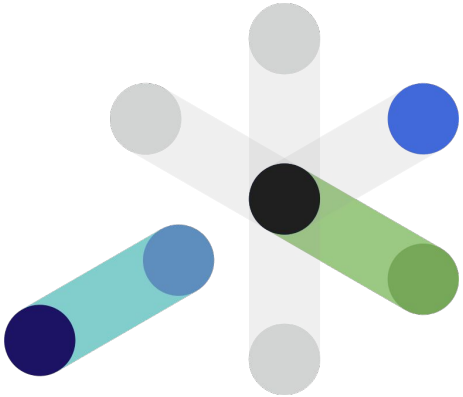
How to get started with data science

srijan:

Step 5:

Be practical

- Start with ML competitions on Kaggle, Hackerearth, Machine Hack, BitGrit
- Get familiar with approaches by reading others approach to a problem (source: Kaggle kernels)
- Follow DFS
- Bonus (earn some good ranks in competitions as a solo member)



Specialization domains

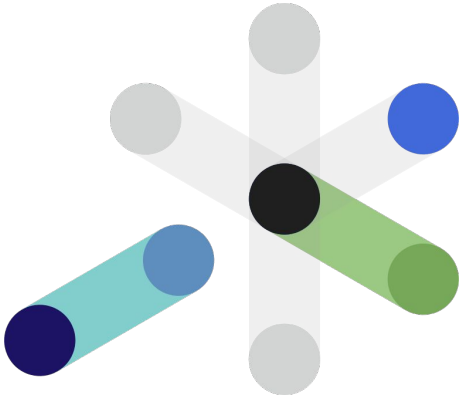
srijan:

Computer Vision (CV)

Way to make machines able to understand and interpret the visual world.

Key thing to focus

- Focus more on image processing (traditional methods)
- Learn complete history of object detection, recognition and segmentation models from traditional methods to current SOTA methods
- Focus more on scalability and simplicity
- Recommended sources: OpenCV documentation, Stanford CS231n, Coursera deep learning specialization



Specialization domains

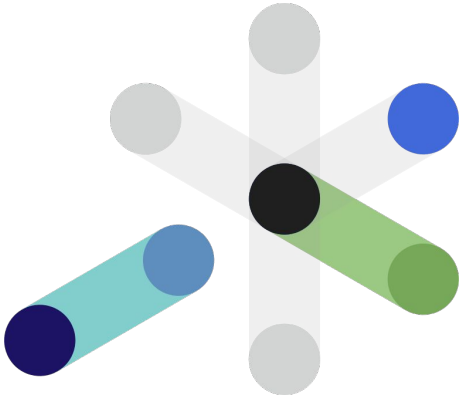
srijan:

Natural Language Processing (NLP)

Way to make machines able to understand and interpret the textual data.

Key thing to focus

- Focus more on classical approaches (TF-IDF, POS, RegExp, text mining, etc)
- Learn complete history of language models from traditional methods to current SOTA methods (RNN, LSTM, Attention, Transformers, etc)
- Focus more on scalability and simplicity



How to prepare for data science interviews

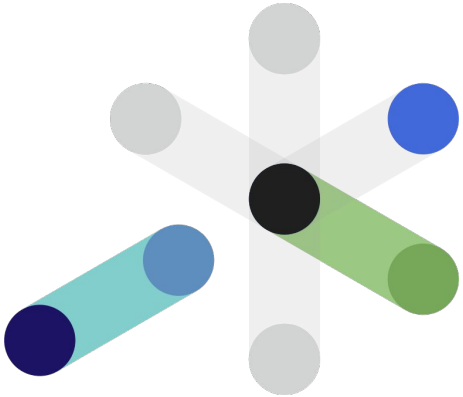
srijan:

Process

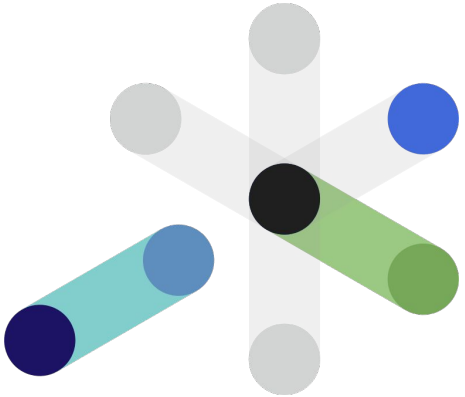
- Online or offline problem statements
  - EDA's
  - Preprocessing steps
  - Modeling techniques
  - Validation Strategies
  - Error Analysis

- Interviewer: Senior data scientist
  - Difficulty level: Medium
  - Resume briefings, solution strategies, why?
  - Understanding of ML algorithms, bagging, boosting, innerworkings, shortcomings, your selection strategy for picking any algorithm, what worked and what didn't ?

- Interviewer: Hiring manager, Practise heads, Principal data scientist
  - Difficulty level: Hard
  - Advanced ML algorithms, inner workings, advantages, disadvantages related to past projects
  - Advanced statistics: Statistics linking with real life
  - Guesstimate problems: Try to be precise rather than totally accurate
  - Case scenarios: Business problems, solutions, strategy, analytical thinking, problem formulation, assumptions



Any Questions ?



Thank You

srijan:

Head Offices

2430 Highway 34
Building B, Suite 22
Manasquan, NJ 08736, USA

8D Vandana Building, Tolstoy Marg,
New Delhi 110001, INDIA

email: business@srijan.net
web: srijan.net
twitter: @srijan