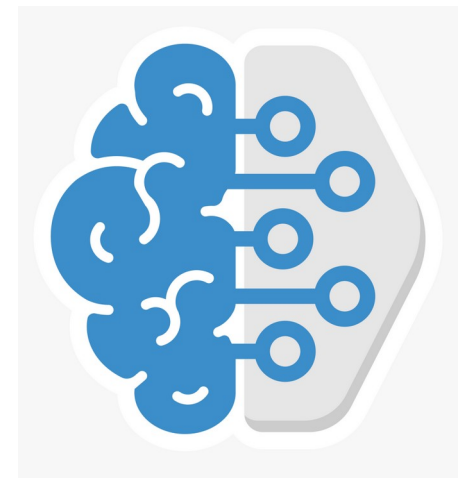


# Machine Learning

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



# Objectives

- Machine Learning Definition and
- Relation with Data Science

# Machine Learning

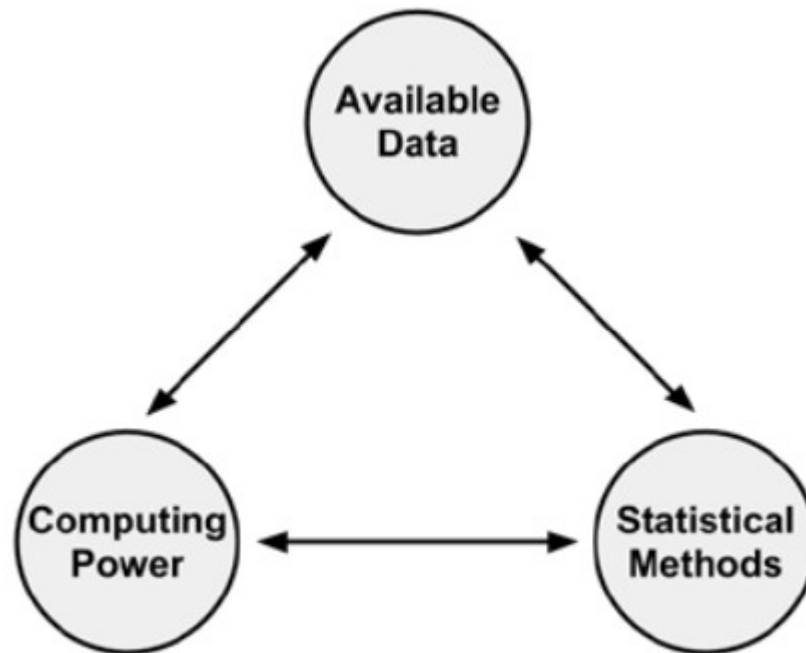
- Machine learning is an application of **artificial intelligence** (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for 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

- 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.

# 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

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 मॉडेल्स 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



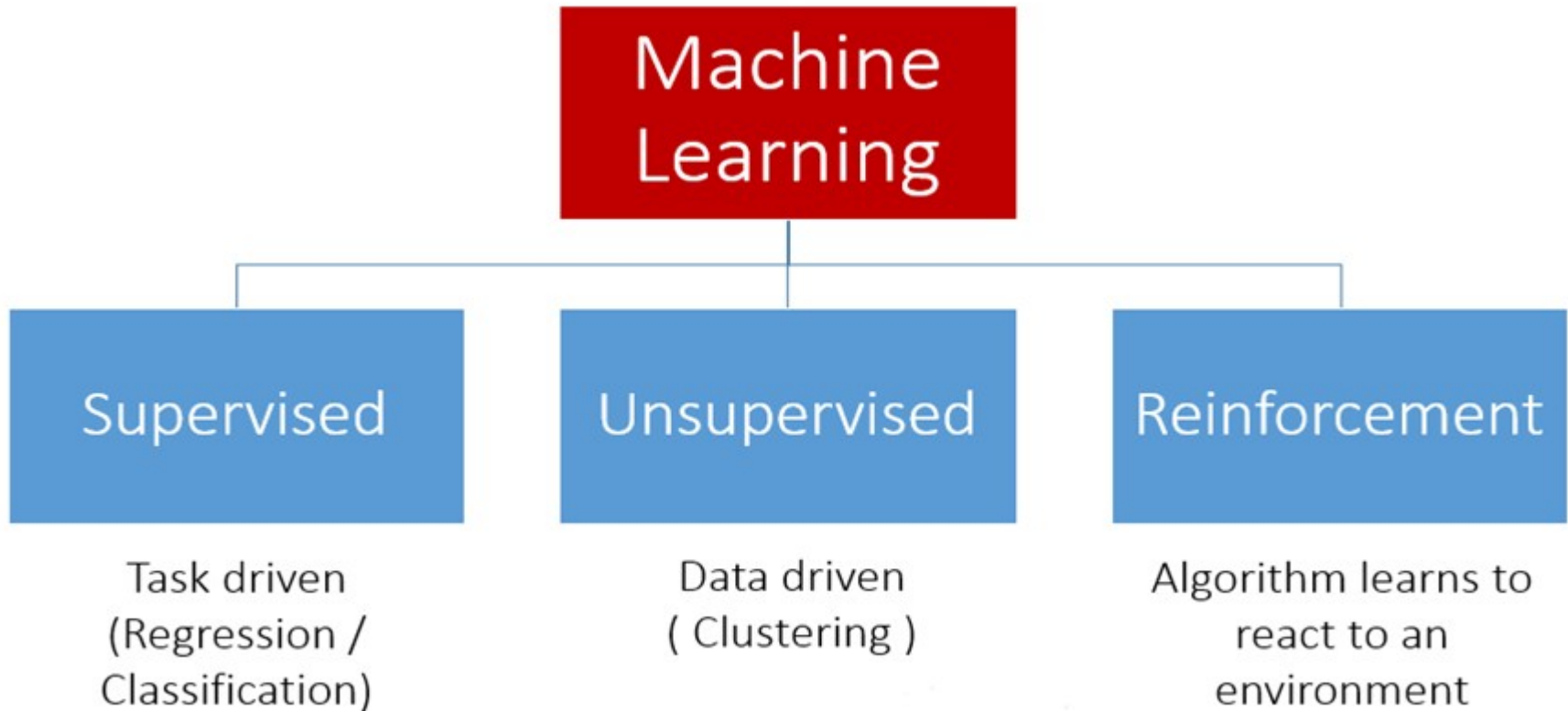
# Recognizing patterns

- 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.

# Types of Machine Learning

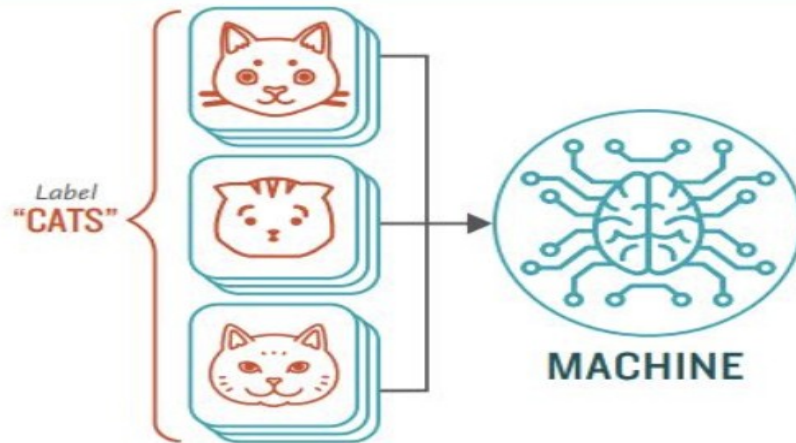


# Supervised Machine Learning

## How **Supervised** Machine Learning Works

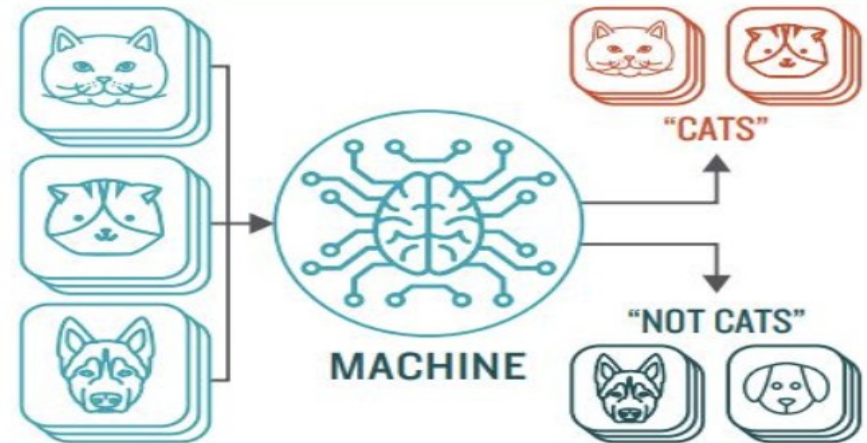
### STEP 1

Provide the machine learning algorithm categorized or "labeled" input and output data from to learn

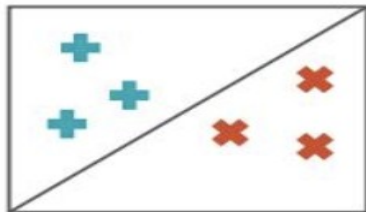


### STEP 2

Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm

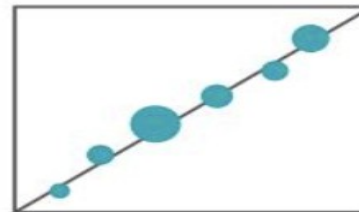


## TYPES OF PROBLEMS TO WHICH IT'S SUITED



### CLASSIFICATION

Sorting items into categories



### REGRESSION

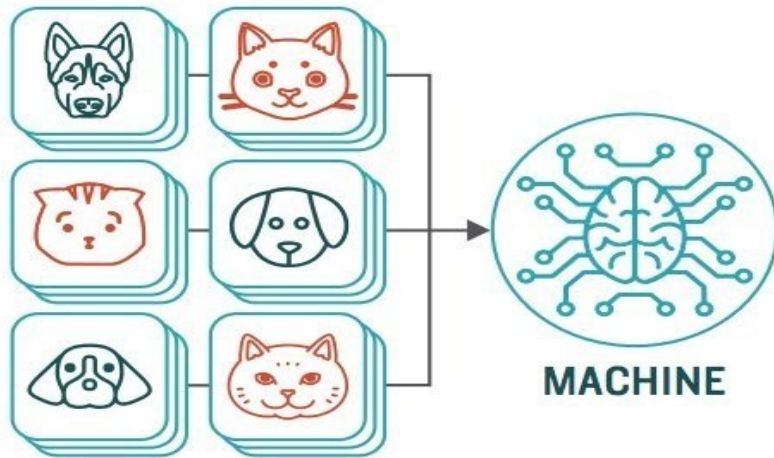
Identifying real values (dollars, weight, etc.)

# Unsupervised Machine Learning

## How **Unsupervised** Machine Learning Works

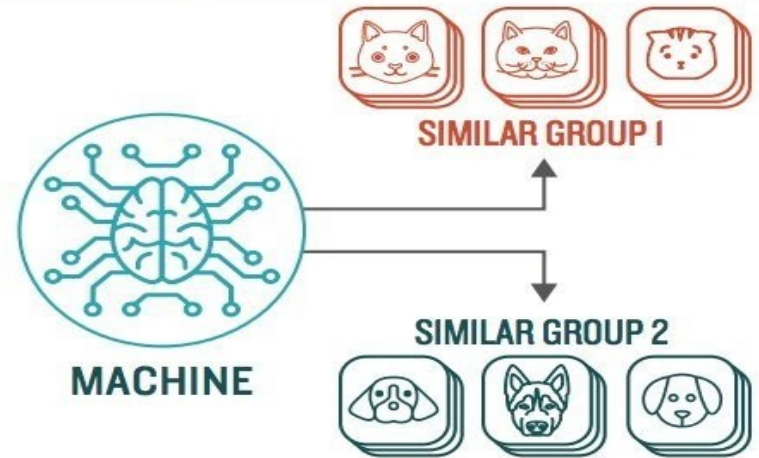
### STEP 1

Provide the machine learning algorithm uncategorized, unlabeled input data to see what patterns it finds



### STEP 2

Observe and learn from the patterns the machine identifies

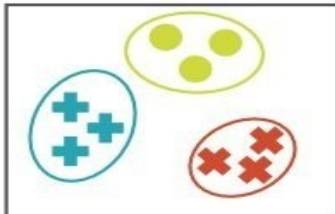


## TYPES OF PROBLEMS TO WHICH IT'S SUITED

### CLUSTERING

Identifying similarities in groups

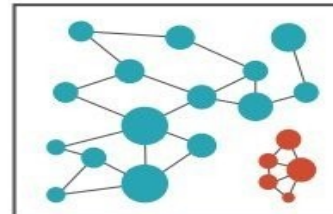
*For Example:* Are there patterns in the data to indicate certain patients will respond better to this treatment than others?



### ANOMALY DETECTION

Identifying abnormalities in data

*For Example:* Is a hacker intruding in our network?



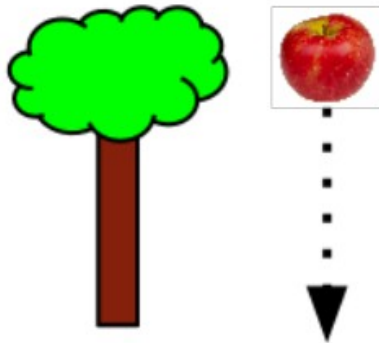
# Training a dataset

- 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.



# Training a dataset

Observations → Data → Model



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

$$g = 9.8 \text{ m/s}^2$$

# 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.

# Data Scientist Skillset



# Machine Learning Skillset



# Summary

## Skills Needed for Data Scientists

- Statistics
- Data mining and cleaning
- Data visualization
- Unstructured data management techniques
- Programming languages such as R and Python
- Understand SQL databases
- Use big data tools like Hadoop, Hive and Pig

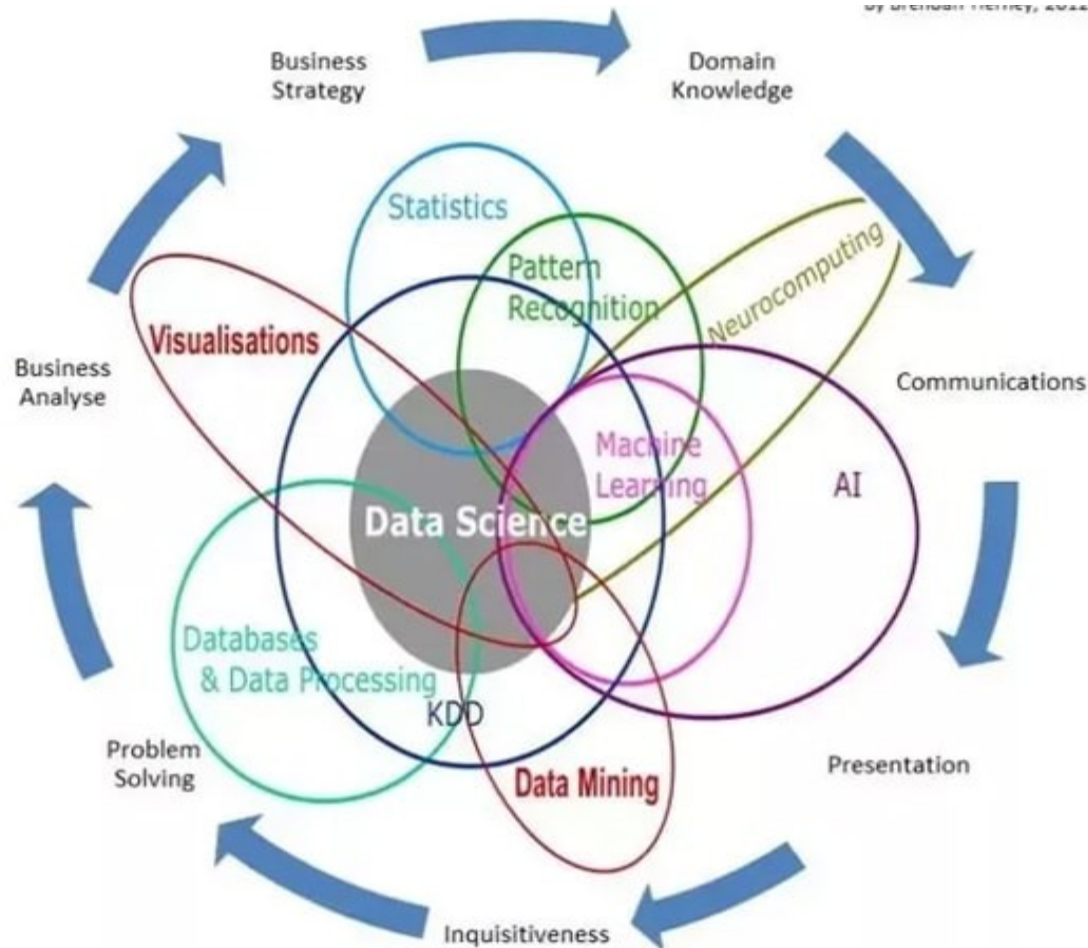
## Skills Needed for Machine Learning Engineers

- Computer science fundamentals
- Statistical modeling
- Data evaluation and modeling
- Understanding and application of algorithms
- Natural language processing
- Data architecture design
- Text representation techniques

# Data Science vs. Machine Learning

- Because data science is a broad term for multiple disciplines, machine learning fits within data science.
- Machine learning uses various techniques, such as regression and supervised clustering. On the other hand, the data in data science may or may not evolve from a machine or a mechanical process.
- The main difference between the two is that data science as a broader term not only focuses on algorithms and statistics but also takes care of the entire data processing methodology.

# Data Science Disciplines



# Applications of ML in Data Science

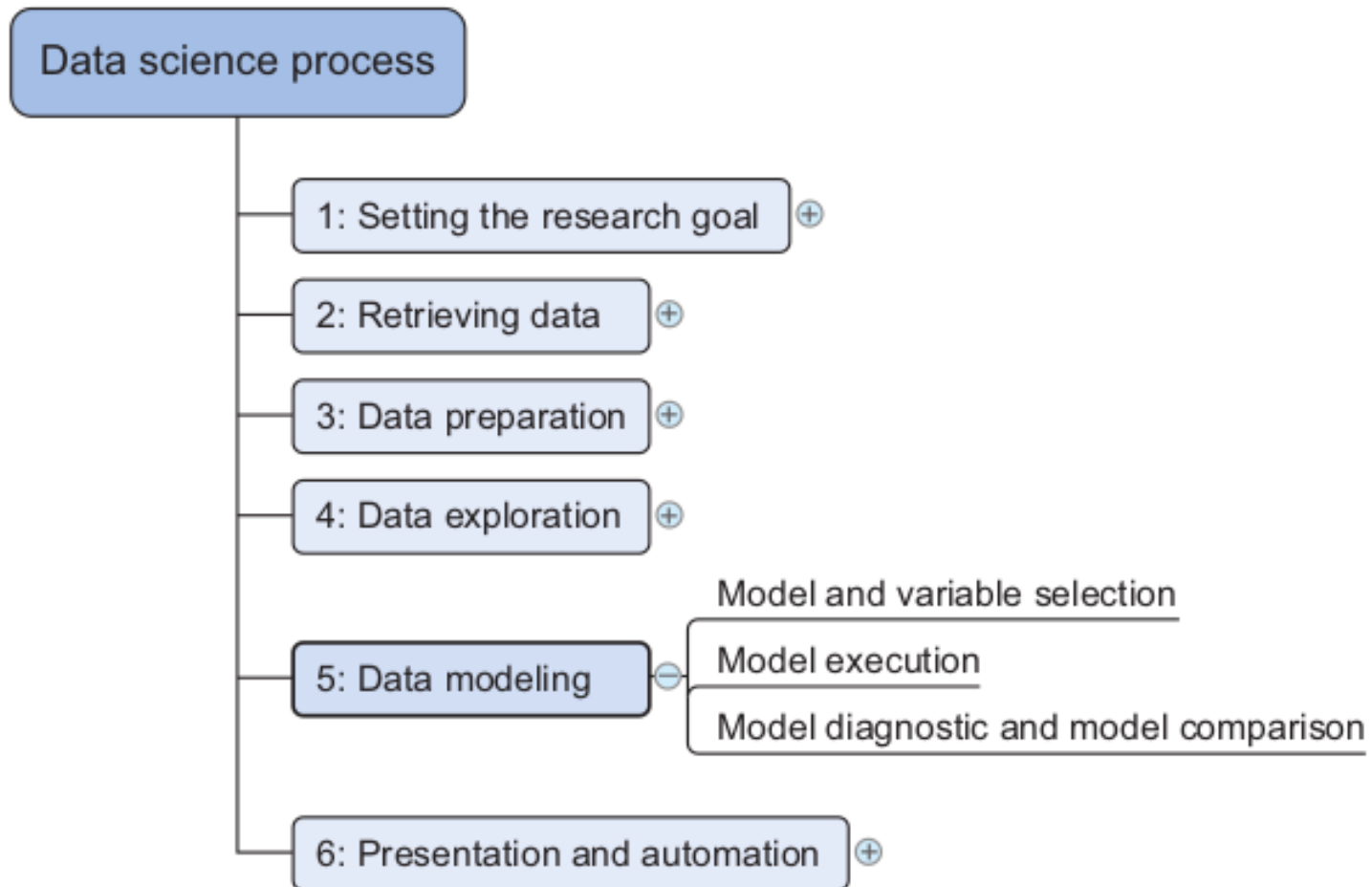
- Regression and classification are of primary importance to a data scientist. To achieve these goals, one of the main tools a data scientist uses is machine learning. The uses for regression and automatic classification are wide ranging, such as the following:
  - Finding oil fields, gold mines, or archeological sites based on existing sites (classification and regression)
  - Finding place names or persons in text (classification)
  - Identifying people based on pictures or voice recordings (classification)
  - Recognizing birds based on their whistle (classification)

# Applications of ML in Data Science

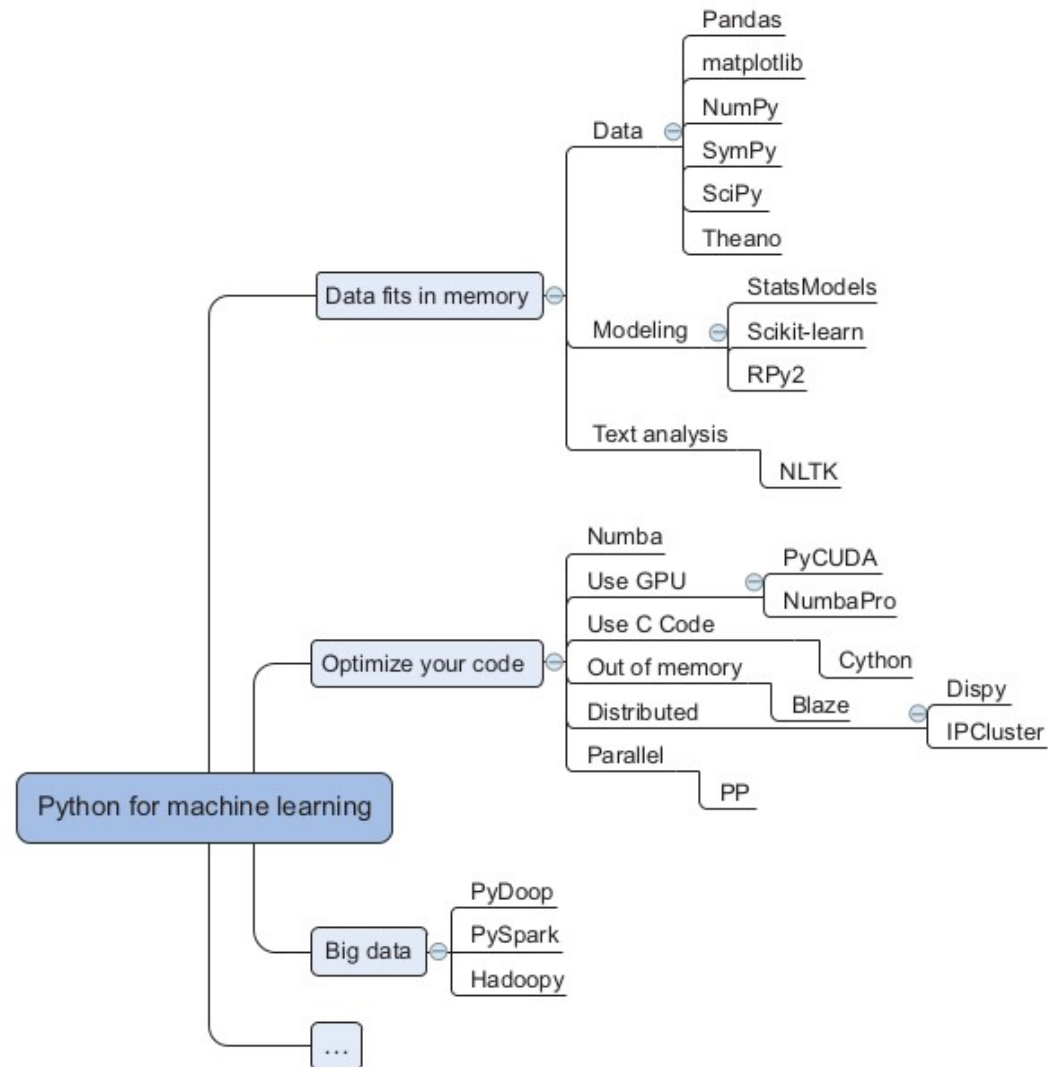
- Identifying profitable customers (regression and classification)
- Proactively identifying car parts that are likely to fail (regression)
- Identifying tumors and diseases (classification)
- Predicting the amount of money a person will spend on product X (regression)
- Predicting the number of eruptions of a volcano in a period (regression)
- Predicting your company's yearly revenue (regression)
- Predicting which team will win the Champions League in soccer (classification)



# Applications of ML in Data Science



# Python for Machine Learning



# Thank you

*This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License*



@mitu\_skillologies



/MITuSkillologies



@mitu\_group



/company/mitu-  
skillologies



MITUSkillologies

## Web Resources

<https://mitu.co.in>

<http://tusharkute.com>

[contact@mitu.co.in](mailto:contact@mitu.co.in)

[tushar@tusharkute.com](mailto:tushar@tusharkute.com)