

Enjoy Data Science

Data Science 101

Schun Uechi

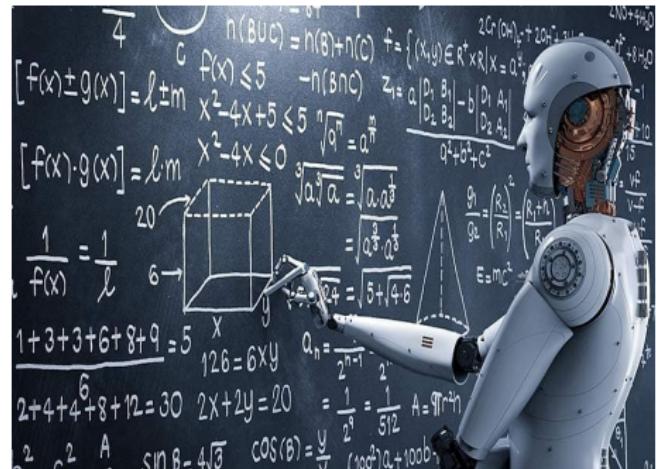
Contents and Prerequisites

- Introduction to Data Science
 - Relation between Data Science and Machine Learning
 - Methodologies of Data Science and Mainstream of Machine Learning Approaches
 - Applications and Examples of Data Science
- Hands on Exercise
- Prerequisites
 - Basic understanding of probability e.g. $100\% = 1.0$.

Data Science

What is Data Science?

Data Science



Data Science



Data Science

Some of the definitions of Data Science.

- Data science is an interdisciplinary academic field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured, and unstructured data¹.
- Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data².
- Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data³.

¹https://en.wikipedia.org/wiki/Data_science

²<https://www.ibm.com/topics/data-science>

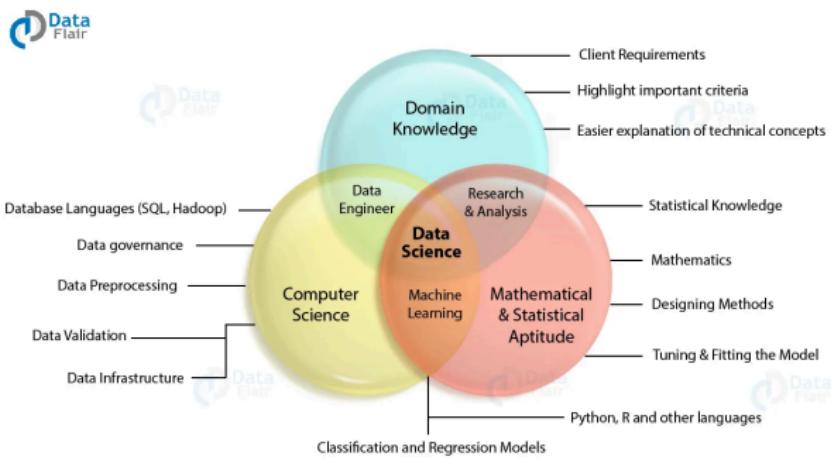
³<https://www.datarobot.com/wiki/data-science/>

What is Data Science?

- Data Science is a “data” oriented science which finds a relation of given data

What is Data Science?

- Data Scientist is a solution provider with synthesizing multidiscipline fields.
- Leverage Artificial Intelligence techniques to find a solution.



Main Stream of Machine Learning Approaches

- Supervised Learning
 - The goal of the task is taught by humans in a supervised manner.
- Unsupervised Learning
 - Humans explore data and seek what can be associated with it.
- Reinforcement Learning
 - Human teaches how to do the task and let the machine do it semi/automatical manner.

What is Data Science?

- Data Science has a lot of methodologies from science
- Let's review an example from science methodology and make a connection with Data Science

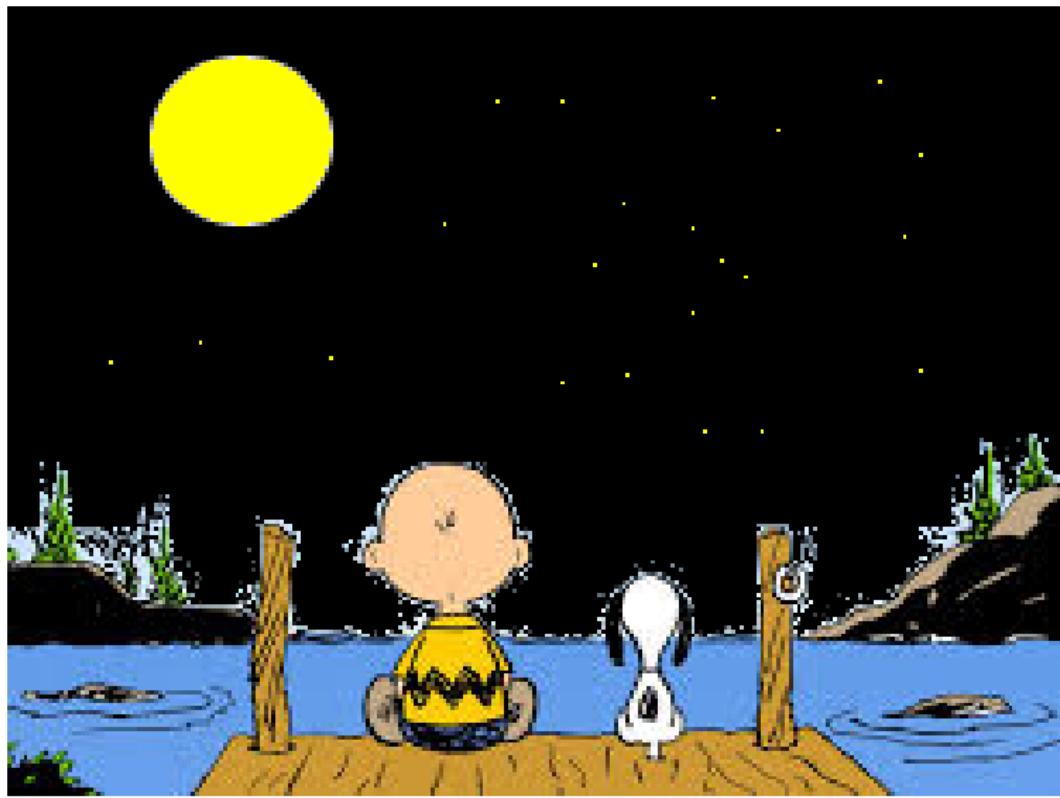
Review Science Methodology

General science approach

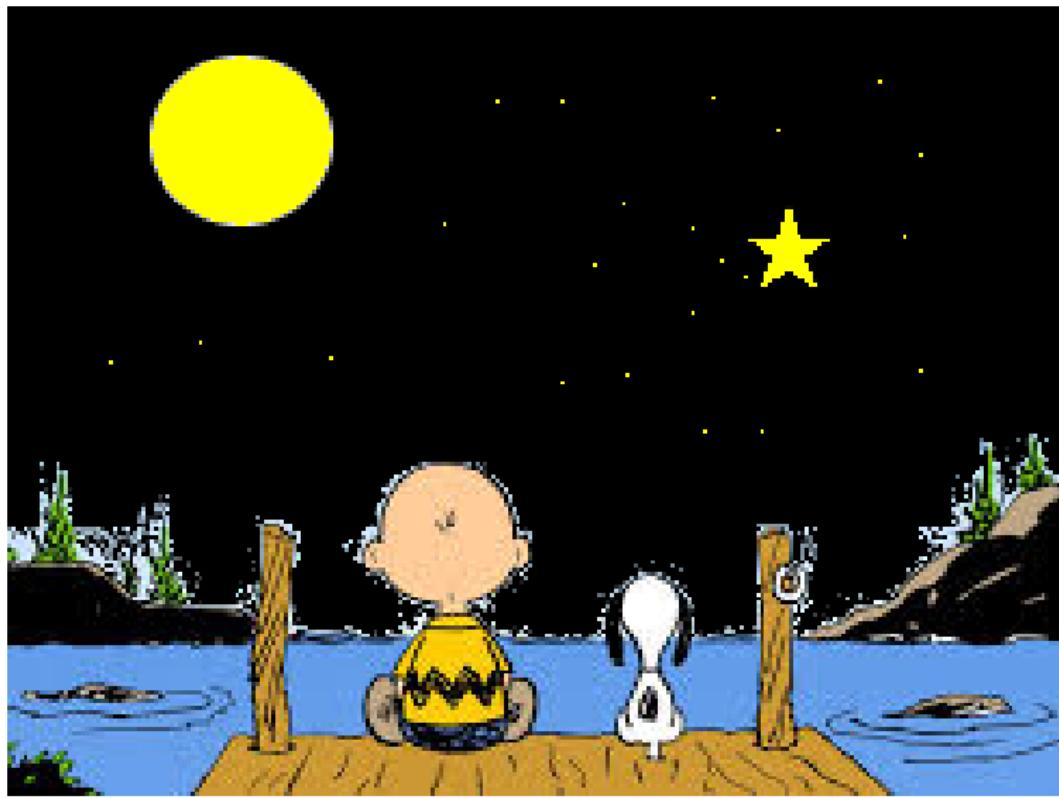
- ① Observation
 - Find or encounter a phenomenon
- ② Hypothesis
 - Make a hypothesis to explain the phenomenon
- ③ Exploration and Experimentation
 - Find example or counterexample and conduct experimentation
- ④ Validation
 - Validate your hypothesis and make a conclusion

Observation

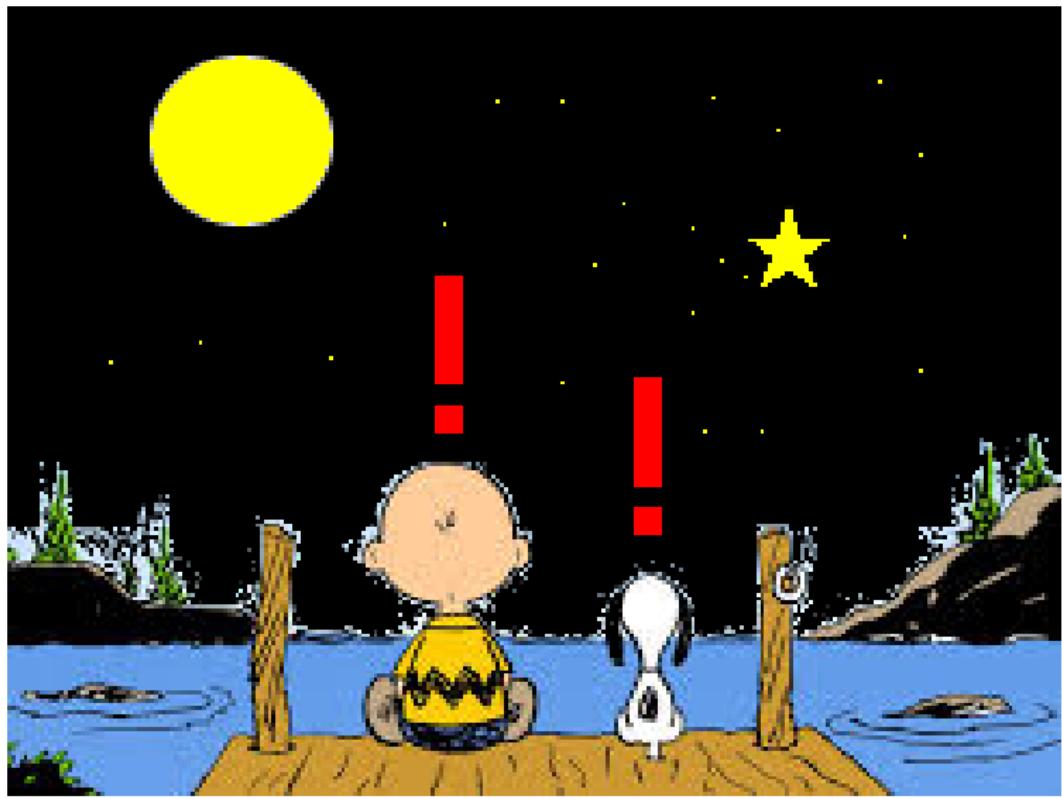
Example: A.C. 1231 in Japan



Example: A.C. 1231 in Japan



Example: A.C. 1231 in Japan



Hypothesis

Example: A.C. 1231 in Japan



Exploration and Experimentation

Example: A.C. 1231 in Japan



2-9

Validation

Oldest Record of Supernova in Japan

It was written in Chinese recode!
Not a sign of omen!!



八日、乙未、霜凝、天晴、北山雪白、客星事、依不審問泰
倭朝臣、過事如此、曉夕東西之條驚而有餘、客星一昨
日夜前空現候了、出現以後去二日陰雲不見候、其外者
天快晴、速日見候、而此兩日者無引運、天曉見良方候、
曉夕東西出現候之條以外候、
客星出現例、皇極天皇元年秋七月、甲寅、客星八月、
陽成院貞觀十九年正月廿五日丁酉戌時、客星在辟、
見西方、宇多天皇寔平三年三月廿九日、己卯、亥時、客
星在東成星東方、相去一寸所、醍醐天皇延長八年五月
以後七月以前、客星入羽林中、一條院宣弘三年四月二
日、癸酉、夜以降驕官中有大客星、如焚惑、光明動耀、
連夜正見南方、或云、驕闊照軍星變本體增光暉、後冷
泉院天喜二年四月中旬以後丑時、客星出觜參度、見東
方、辛天闕星、大如歲星、二條院永萬二年四月廿二日、
乙丑、亥時、客星替見大微宮事、高倉院治承五年六月
廿五日、庚午、戌時、客星見北方、近王良星守傅舍星、
午終許心寂房來、事外付滅之由加詞但惡血之充瀉非
嗣雖極難治云々、常隨侍之小婢、依病房危急行南
京、老病之最中失手背之便、

Meigetsuki: Record of astronomical phenomena in Japanese document by Fujiwarano Teika in 11th century.

Compare Science Methodology and Data Science Approach

- Observation
 - Observe phenomena or signal that people want to know the meaning of.
- Hypothesis
 - Is it sign of omen?
- Experimentation and/or Experimentation
 - Find record which is related to the phenomena.
- Validate and/or conclusion
 - Found it was not a sign of omen.

Compare Science Methodology and Data Science Approach

	Science	Data Science
Observation	Find phenomenon	Find data set
Hypothesis	Explaine the phenomenol	Extract information
Experimentation	Field work and experimentation	Machine Learning model etc.
Validation	Validate with the phenomenon	Validate with the data set



Q&A and Hands on Exercise

References

- Code: Hands on Exercise
 - ① <https://github.com/quantumhara/lecture>
- Citations: Books of Data Science for Business
 - ① https://www.researchgate.net/publication/256438799_Data_Science_for_Business
 - ② https://book.akij.net/eBooks/2018/May/5aef50939a868/Data_Science_for_Bus.pdf