

Machine Learning for Data Science & Artificial Intelligence With Python

Introduction to Data Science

H.M. Samadhi Chathuranga Rathnayake

M.Sc in CS (SU), PG.Dip in SML (Othm), PG.Dip in HRM (LRN), B.Sc (Hons) in IS (UOC), B.Eng (Hons) in SE (LMU), P. Dip EP & SBO (ABE), Dip SE, Dip IT, Dip IT & E-Com, Dip B.Mgt, Dip HRM, Dip Eng

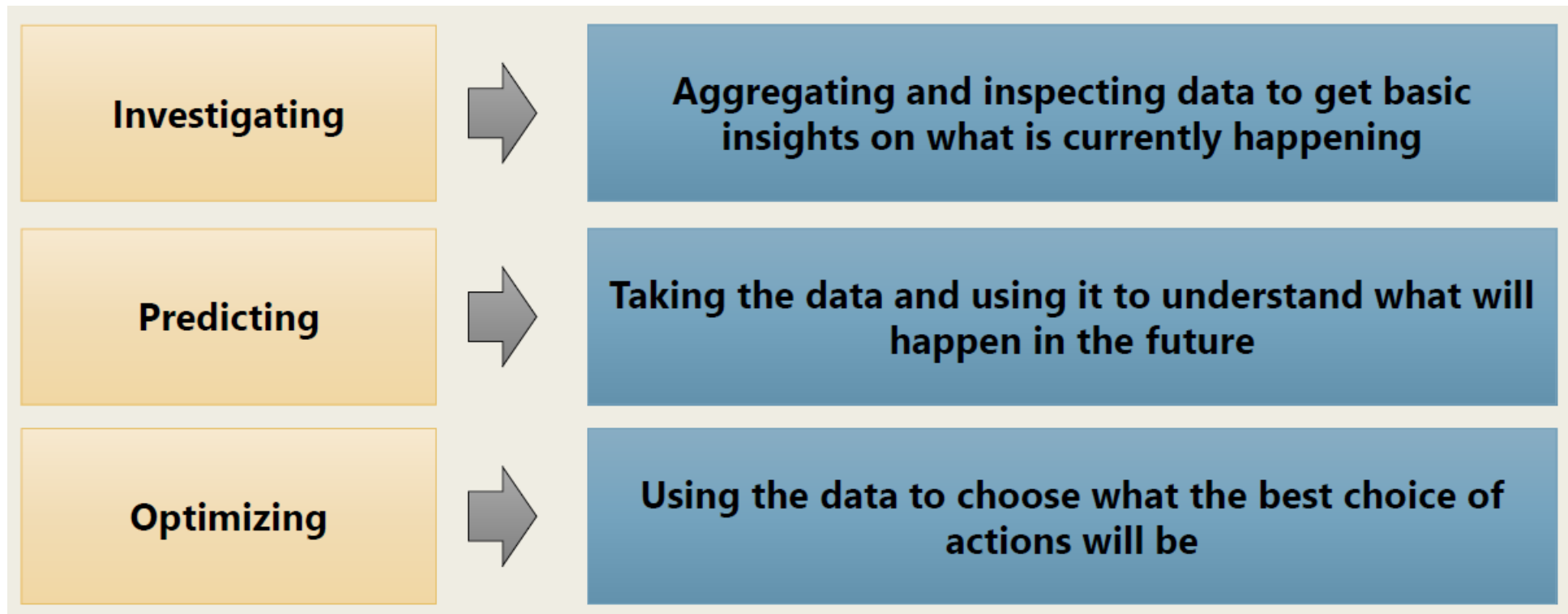
What is Data Science?

- Data science is an interdisciplinary field that uses,
 - Scientific methods
 - Processes
 - Algorithms
 - Systems
- To extract knowledge and insights from data in various forms, both structured and unstructured.



What is Data Science?

- Data science tends to fall into three broad categories :



What is Data Science?

If you have two products, which is selling better? How many people are buying both?

How frequently do customers order?




Investigating

How are sales changing each month?

What is Data Science?

How likely is a customer going to come back?

How will a customer respond to a sale?



Predicting

Is revenue going to increase over time?

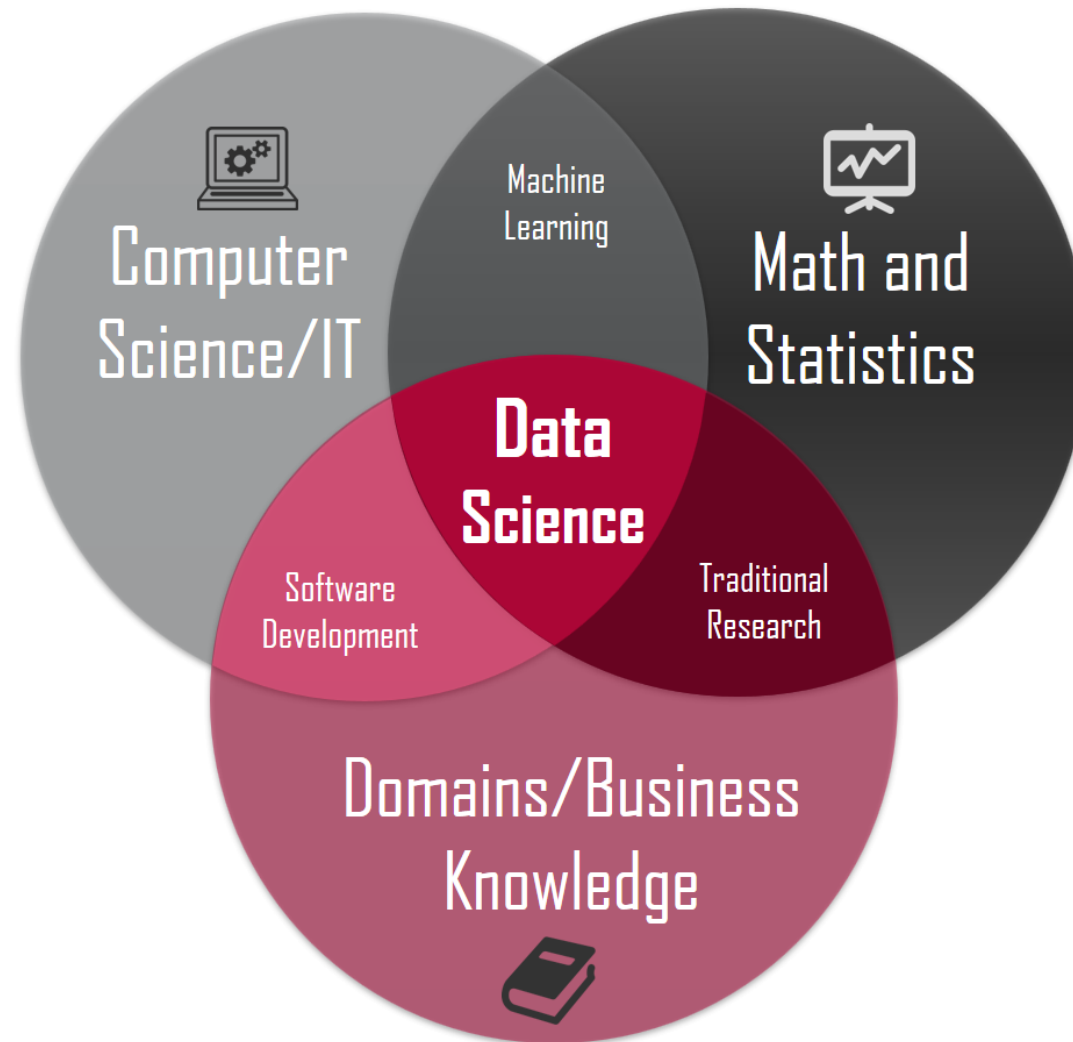
What is Data Science?

How much inventory should be held, and when should you reorder?

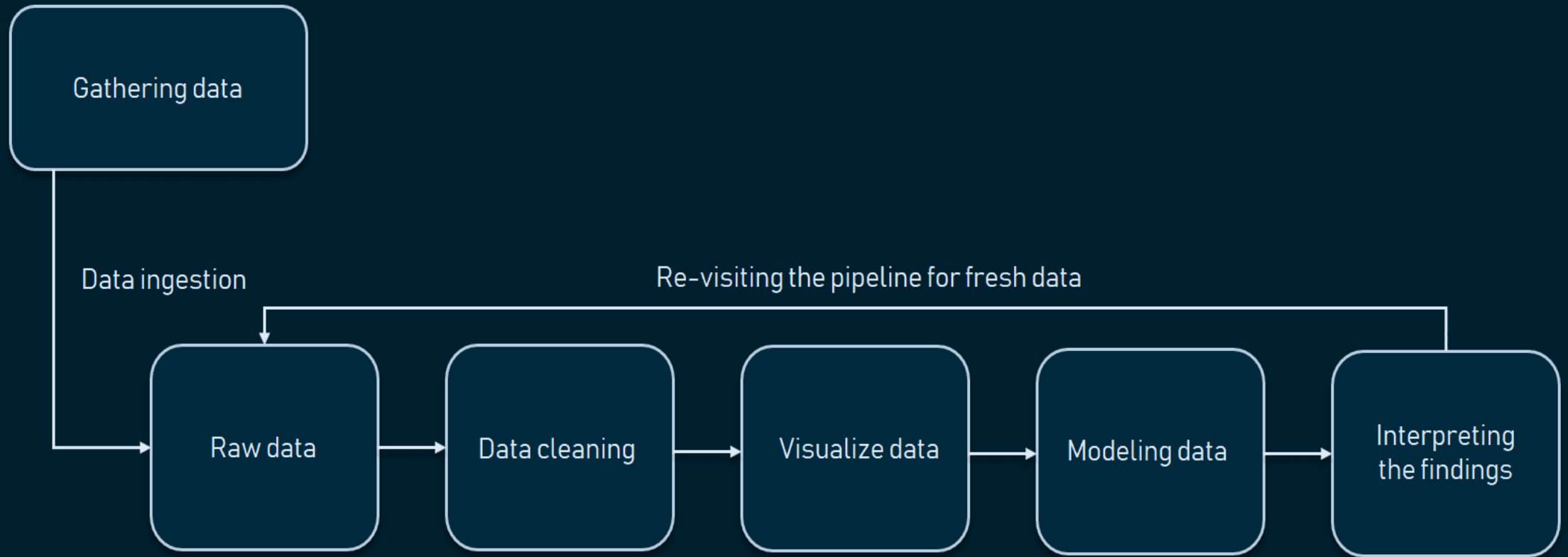


What price should you set for a product? When should it go on sale?

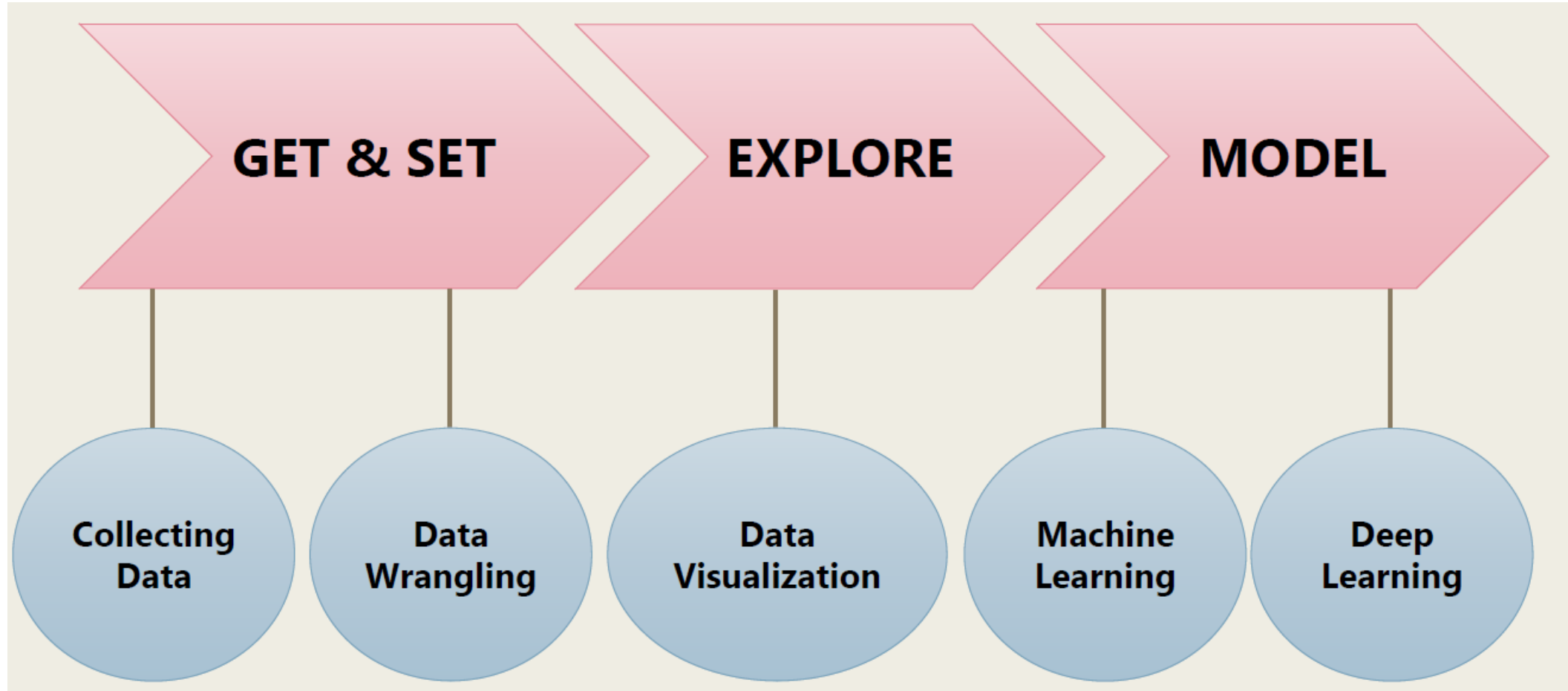
Skills of a Data Scientist



Data Science Pipeline



Data Science Pipeline



Is Data Science A New Thing?

The term “data science” has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. However, the definition was still in flux.

In the 1990s, the term "Data Mining" was introduced, but data mining is the evolution of a sector with an extensive history. Early techniques of identifying patterns in data include Bayes theorem (1700s), and the evolution of regression(1800s).

Data science is a broad field that includes the processes of capturing of data, analyzing, and deriving insights from it. On the other hand, data mining is mainly about finding useful information in a dataset and utilizing that information to uncover hidden patterns.

Why Now?

- Increasing the sources of data
- Technological development
- Theoretical & conceptual development
- Development of the skills