

# DATA SCIENCE WITH MACHINE LEARNING

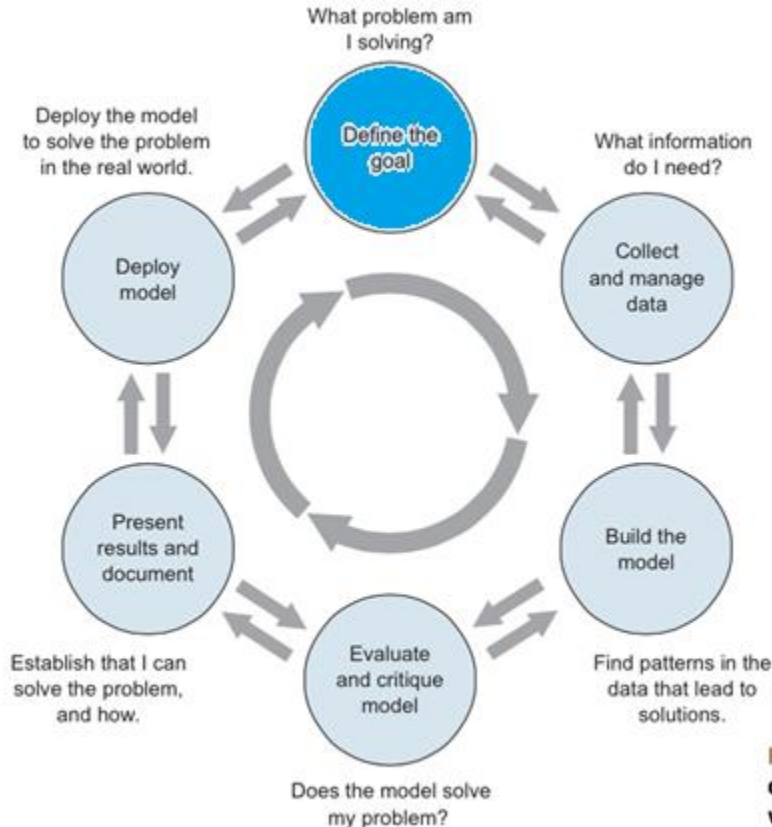
**A JOURNEY FROM DATA TO DECISIONS**

**ZAHID HUSSAIN**

**SR. LECTURER DHA SUFFA UNIVERSITY**

**DEPARTMENT OF COMPUTER SCIENCE**

# General Workflow

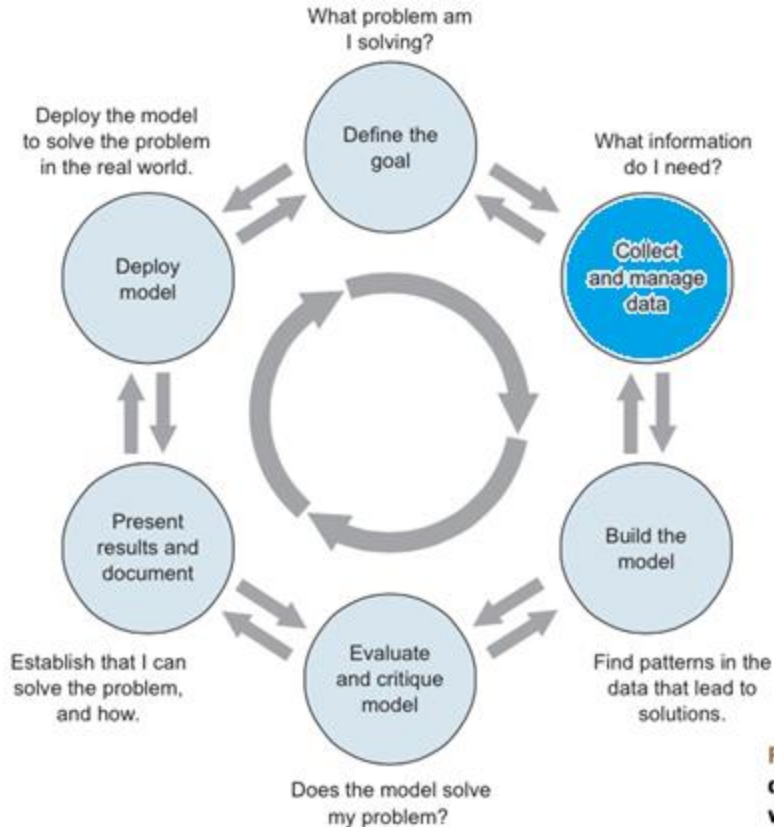


## Defining the Goal

- What is the question/problem?
- Who wants to answer/solve it?
- What do they know/do now?
- How well can we expect to answer/solve it?
- How well do they want us to answer/solve it?

**Figure 1.1** The lifecycle of a data science project: loops within loops

# General Workflow

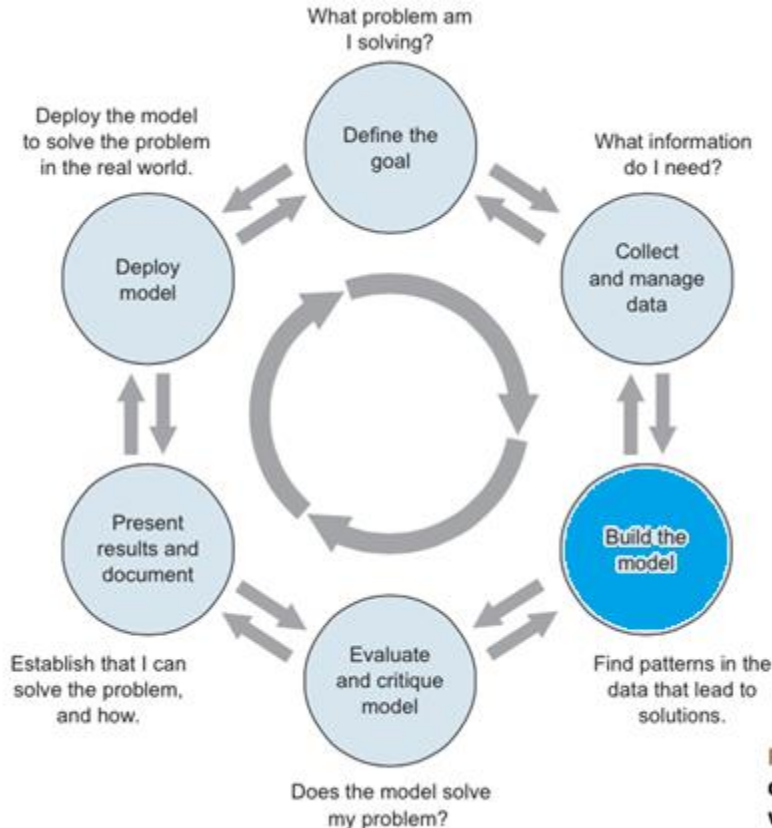


## Data Collection and Management

- What data is available?
- Is it good enough?
- Is it enough?
- What are sensible measurements to derive from this data? Units, transformations, rates, ratios, etc.

**Figure 1.1** The lifecycle of a data science project: loops within loops

# General Workflow

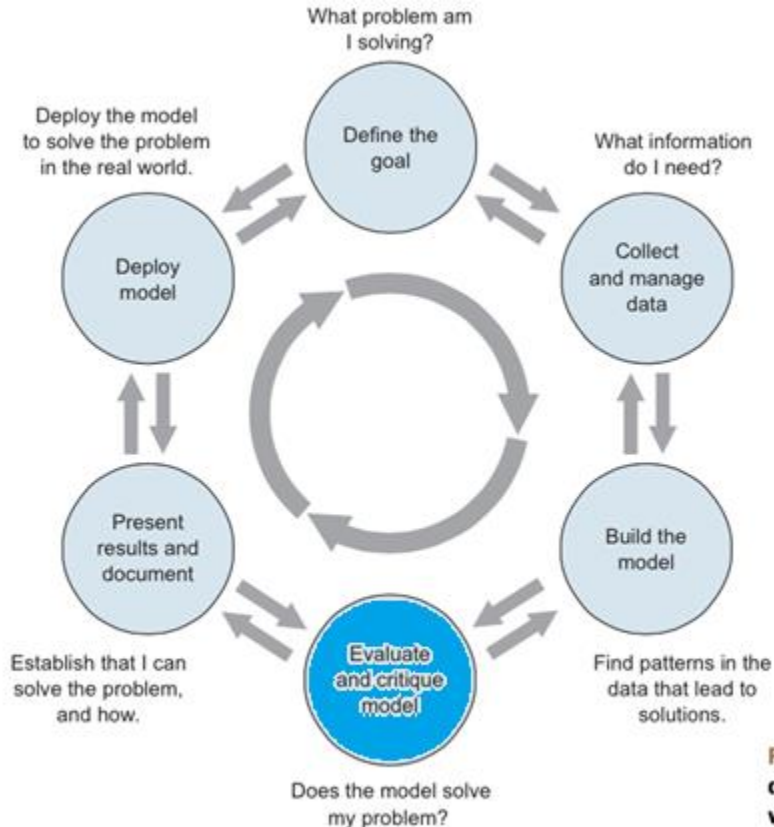


## Modeling

- What kind of problem is it? E.g., classification, clustering, regression, etc.
- What kind of model should I use?
- Do I have enough data for it?
- Does it really answer the question?

**Figure 1.1** The lifecycle of a data science project: loops within loops

# General Workflow

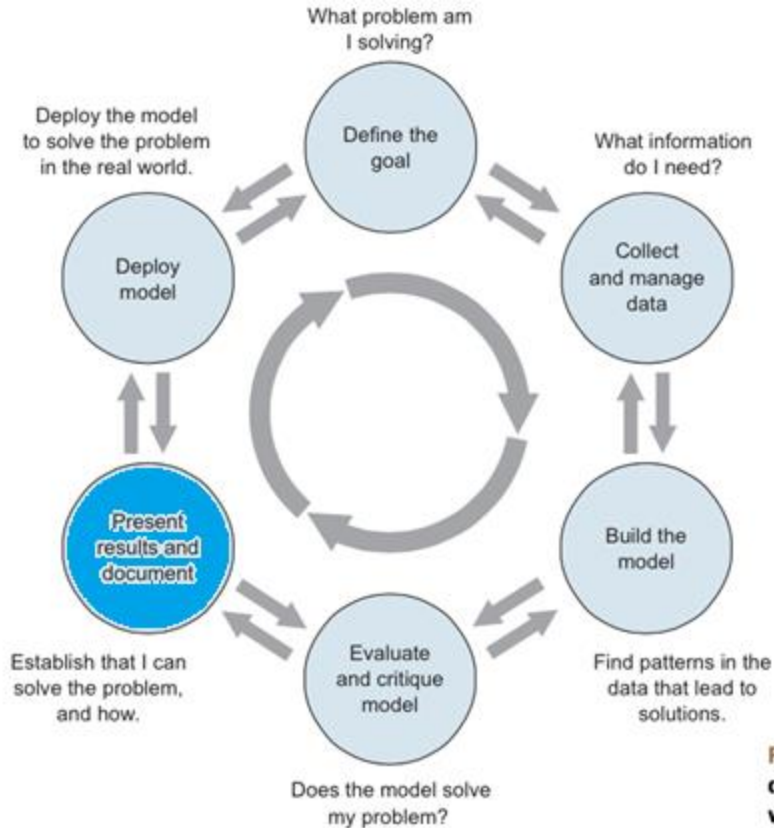


## Model Evaluation

- Did it work? How well?
- Can I interpret the model?
- What have I learned?

**Figure 1.1** The lifecycle of a data science project: loops within loops

# General Workflow

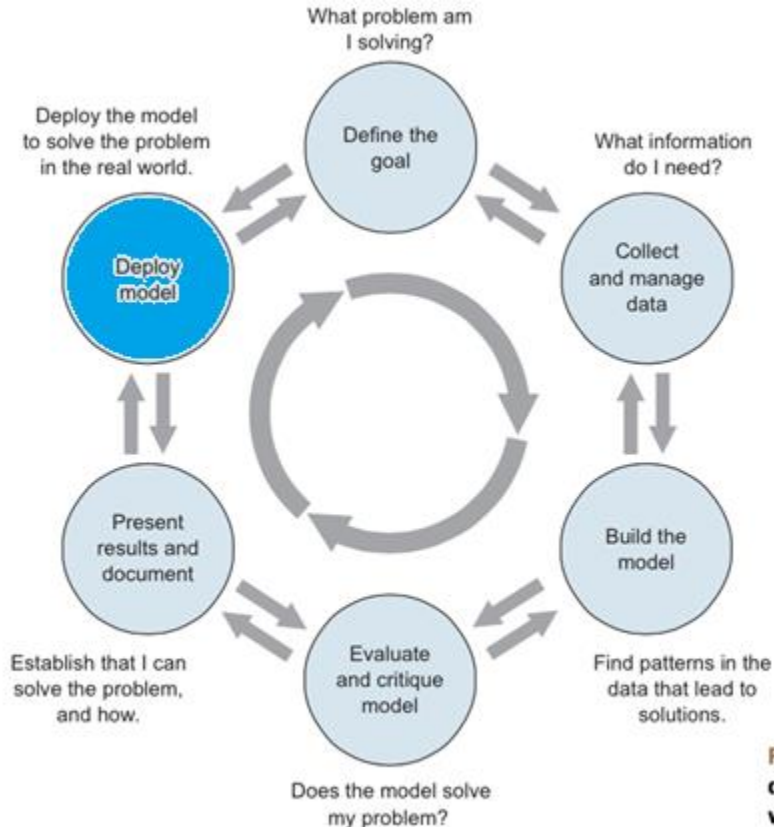


## Presentation

- Again, what are the measurements that tell the real story?
- How can I describe and visualize them effectively?

**Figure 1.1** The lifecycle of a data science project: loops within loops

# General Workflow



## Deployment

- Where will it be hosted?
- Who will use it?
- Who will maintain it?

**Figure 1.1** The lifecycle of a data science project: loops within loops