



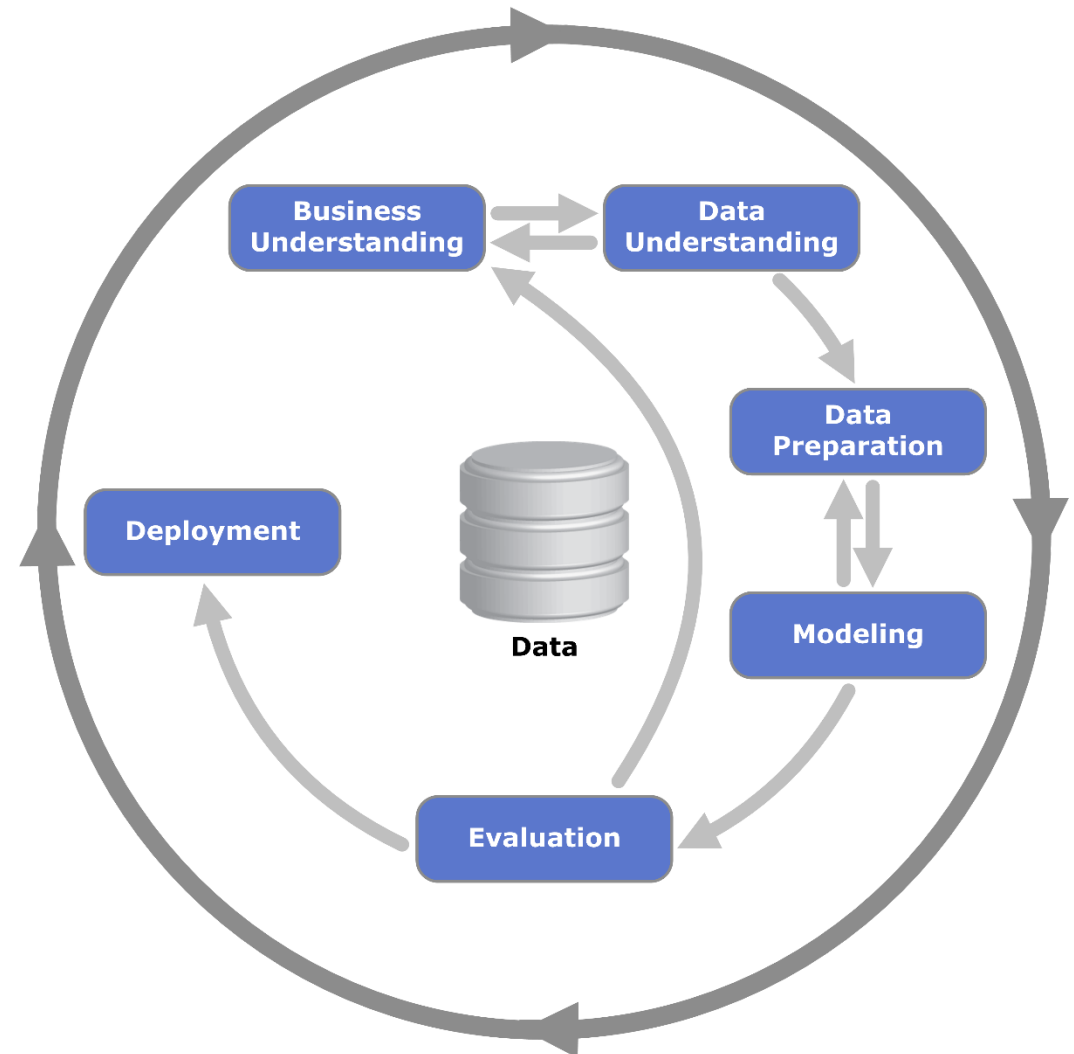
CRISP-DM

Knowledge Discovery Process

Group F

CRISP-DM is the most used methodology for Data Science Projects.

CRoss
Industry
Standard
Process
for
Data
Mining

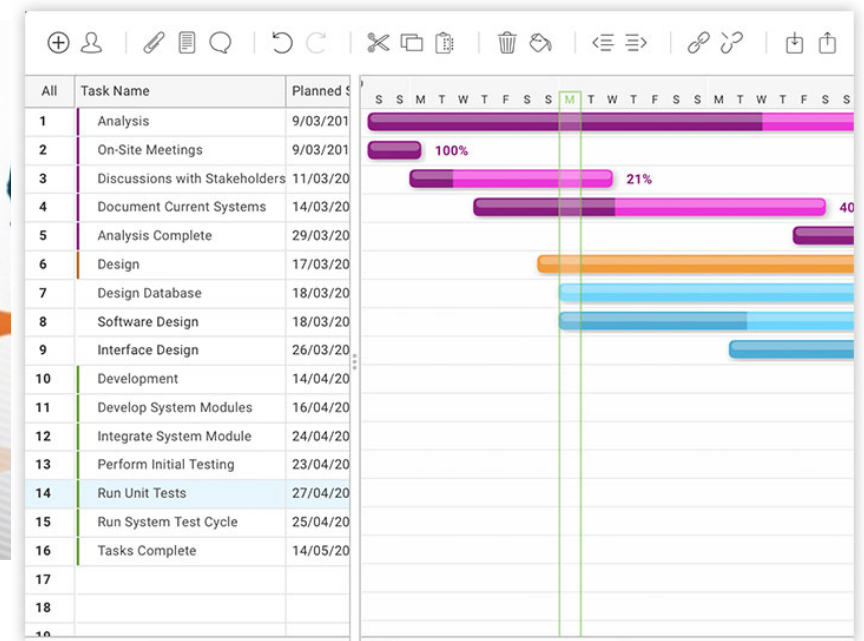


We are going to show you the key things
that you need to learn



Business understanding phase is all about translation to analytical form.

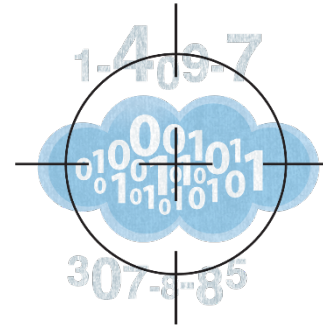
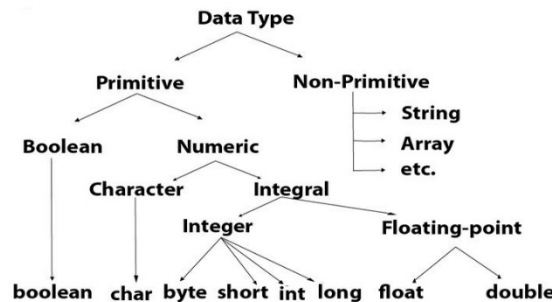
- Understand what the real business problem they want you to solve
- Look for the key players in the game (stakeholders)
- Define the success criteria (KPI)
- Make a project plan



Data Understanding is the dirty job that no one wants to do



shutterstock.com • 477457132



Collect the data:

- What is the right data
- Sources of the data
- Select tables and/fields

Describe the Data:

- Check access to data, volume and complexity
- Understand the meaning of data
- Analyze the relationships

Explore the data:

- Analyze the properties
- Make the first findings
- Get interesting subsets

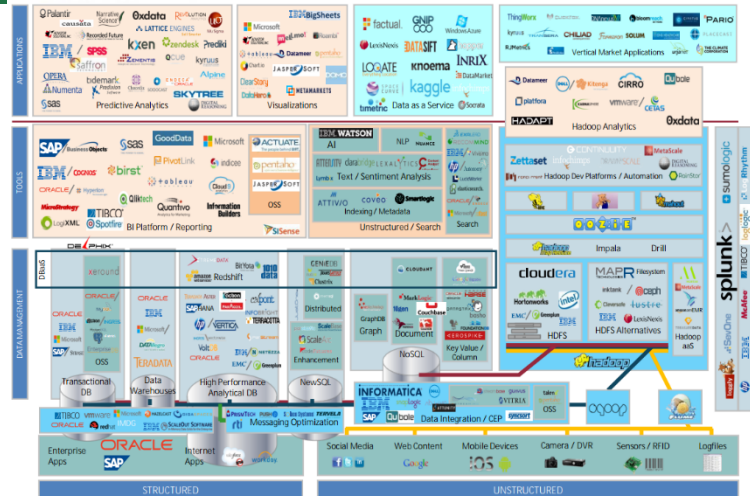
Verify Quality

- Check keys
- Check data issues (missing, blank, typos, spelling)
- Check special values

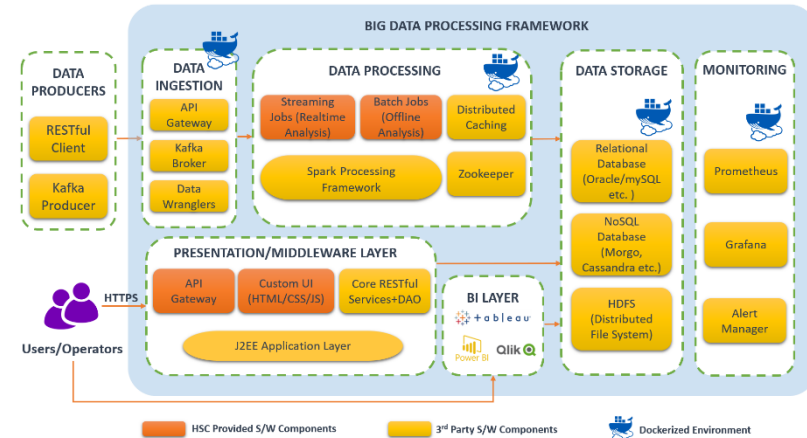
Technology Platform: don't overfit your company with platforms



Underfitted

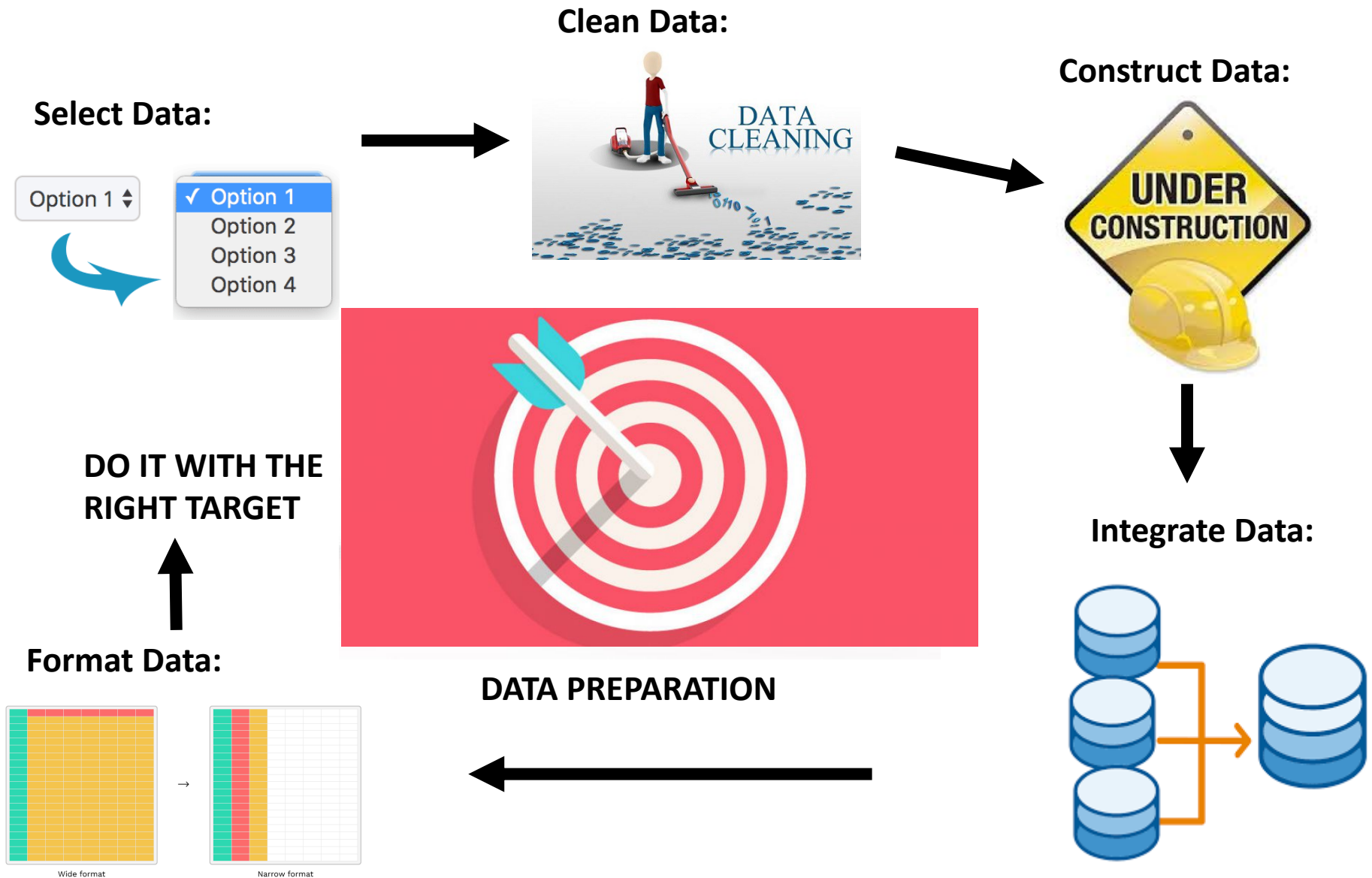


Overfitted



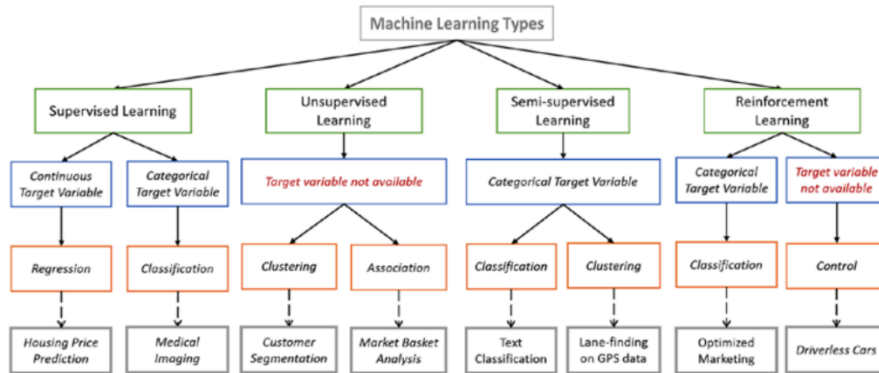
Good Fit/Robust

Data Preparation: this is when the action starts

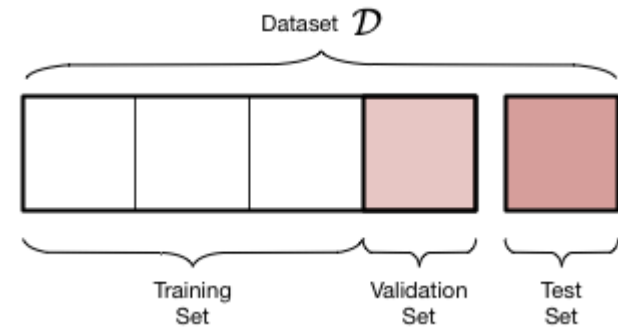


Modeling, start simple and push it forward

Select the fitting models:



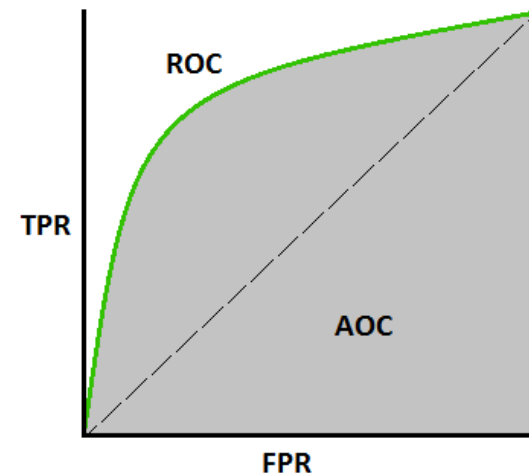
Decide Samples and prepare data:



Optimize the Hyperparameters and run the model:



Assess the model:



Evaluation, time to go back and check on everything you've done before.

Did your model solve the business problem?



Are all the steps necessary?



What are the next steps?



Deployment is to integrate your model into the company structure



Technical Deployment:

- Draw conceptual Model
- Decide on the analytical tool and model integration
- Temporal Planning
- Other systems connections



Business Deployment:

- Decide on the users, key actors for the model
- Produce the full documentation/platforms
- Decide the applicability and threshold of the model
- Plan on monitoring and Maintenance

USE CRISP-DM otherwise...

