6 Principles of a Data Science Mindset

1. Understand the business problem

Q1. Define business goals: discuss with business team and stakeholders to identify the business problems

Q2. Define data science goals by asking right questions: transform the business goals to data science goals

EXAMPLE:

Business goal: Manual project staffing is time consuming and the results are sub-optimal. Right questions to make a data science goal:

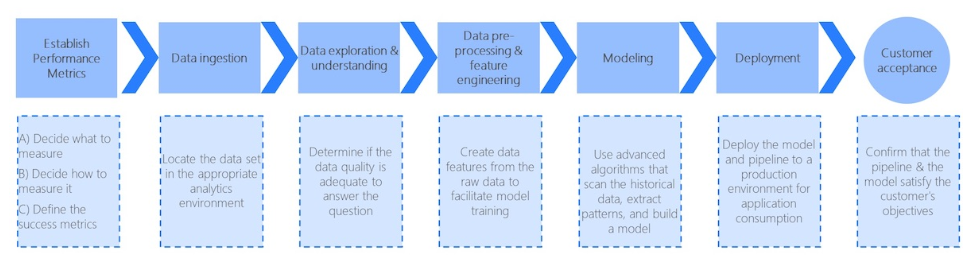
1. Given historical data of each employee, how can we predict staff composition for a new project?
2. How can we compute a staff fitness score for a new project?

1. Establish performance metrics

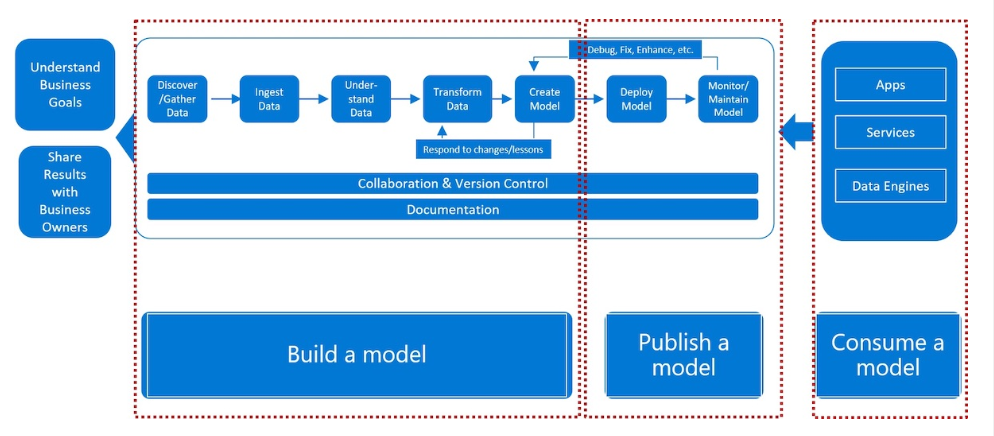
Q1. Define what data to collect to solve the business problem

Q2. Define how to collect and how to make ground truth

Q3. Define the success of your project. Which metric you are trying to monitor?



1. Architect the end-to-end solution



1. Build your toolbox of data science tricks

* Train on a subset much smaller that the whole dataset you have first.
* Reuse knowledge gained from previous projects like data processing/ hyperparameters
* Setup automated alerts that will inform that a specific experiment is over
* Use Jupyter notebooks for quick prototyping. Rewrite the code into python modules/classes once you are satisfied.
* Keep you experiment code in version control system
* Use pre-configured environments in the cloud for data science development
* Have a list of things to do while experiments are running: data collection, cleaning, annotation; reading on new data science topics, experimenting with a new algorithm or a framework

1. Unify the organization’s data science vision

Data science teams and business teams should regularly communicate to share their goals, values and vision

1. Keep humans in the loop

Compare the solution of the machine learning model with the human performance

**Checklist**

1. Define Business Goal
2. Collect data and Define success criteria
3. Prepare folder structure and plan code pipeline
4. Explore and validate data
5. Select a baseline model (through past experience or survey)
6. Quickly Implement baseline and obtain initial result
7. Do error analysis to find problem in baseline
8. Propose model to solve the problem
9. Does obtained result meet success criteria? If no, improve the model from 7 else go to 10.s
10. Present results to business people
11. Are results satisfactory? If no, improve the model from 7 else go to 12.
12. Test the model for different data
13. Deploy and maintain the model
14. Create services, apps, engines